DOCUMENT RESUME

ED 176 987 SE 028 914

AUTHOR	Cross, Judson B.; And Others
TITLE	Applied Mathematics: An Introductory Course.
INSTITUTION	Boston Univ., Mass. Physical Science Group.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	75
GRANT	NSF-GZ-2892
NOTE	360p.; Contains occasional light type
EDRS PRICE	MF01/PC15 Plus Postage.
DESCRIPTORS	*Calculation; College Curriculum; *College Mathematics; Higher Education; Linear Programing; *Mathematical Applications; *Mathematical Concepts; Mathematical Enrichment; Mathematical Vocabulary; *Mathematics; *Mathematics Instruction; Science Education; Set Theory; Trigonometry
IDENTIFIERS	*Functions (Mathematics)

ABSTRACT

The intent of this text is to provide students in a variety of science and technology disciplines with a basic understanding of mathematics commonly used in introductory texts in such disciplines. The first five chapters develop skills needed for efficient numerical calculation. The last five chapters examine the basic properties of elementary functions. Special emphasis is placed on finding analytical expressions from graphical representation of data. (Author/RE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED176987

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY
Mary L. Charles
NSF

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Archives of
HIGHER EDUCATION IN SCIENCE
ATP/SCIP/MIDS

Property of
THE NATIONAL SCIENCE FOUNDATION

GZ-2892
Boston U.
Uri-Haber-Schaim

APPLIED MATHEMATICS

AN INTRODUCTORY COURSE

PHYSICAL SCIENCE GROUP
BOSTON UNIVERSITY

SE 028 914

Copyright © 1974, 1975 by The Trustees of Boston University

Printed in the United States of America

PREFACE

Are there basic needs in applied mathematics that are shared by beginning college students in the social sciences, the natural sciences, and technology?

Several topics come to mind that point to an affirmative answer: presenting and interpreting data, finding analytical expressions for functions from graphs, being familiar with the properties of elementary functions, and being conversant with the language of calculus. By providing for these needs, we enable the students to overcome serious obstacles to the understanding of introductory texts in all these fields. The intention of this text is to serve just this purpose.

The first five chapters develop the skills needed for efficient numerical calculations, emphasizing the consequences of the inherent uncertainties of most numbers used in applications. The topics discussed range from order-of-magnitude estimates through the theory and the use of the slide rule to the fundamentals of the use of computers. (Although the importance of the slide rule is declining because of the growing use of calculators, an understanding of the logarithmic scale is as important as ever.)

The last five chapters examine the basic properties of the elementary functions, including their derivatives and integrals. Special emphasis is placed on finding analytic expressions from graphical representation of data.

The book has been written with an interactive mode of learning in mind. It is suitable for section work where short lectures, discussion of text, and problems can be carried out as needed. Whenever we believed that certain points are best made by having the students tackle them, these points were included in the questions at the end of the section. Thus, the questions form an integral part of the course. Many of the questions can be approached in different ways and thereby present the opportunity for constructive discussion and a means for improving the communicative skills of the students. There are relatively few drill problems. Extra problems of this kind can easily be provided by the instructor.

Because questions are placed after each section, the text may also be used for individual study.

. . . .

This book has its origin as the freshman mathematics course in our Undergraduate Program for Physics-Chemistry Teachers that started in 1970. However, since then it has also been used extensively by students in other fields.

The principal contributors to the preliminary edition were Judson B. Cross, Thomas J. Dillon, Jo Rita Jordan, George Lukas, Leonard T. Nelson, Poul Thomsen, David B. Teague, and myself.

This book constitutes a far-reaching revision of the preliminary edition, including much new material. The revision was done by Judson B. Cross, Robin Esch, Romualdas Skvarcius, and myself.

The revision benefited from the feedback of the following professors who piloted the course: Leonard T. Nelson and Joseph Van Wie at Southwest Minnesota State College, Henry P. Guillotte at Rhode Island College, and Albert G. Starling and David B. Teague at Western Carolina University.

The work was illustrated by George Frigulietti and Myrna S. Goldblat, and produced by Benjamin T. Richards. The bulk of the camera copy was typed by Caroline E. Russell; the typing was completed by Lorraine Perrotta.

The development of this book is supported by a grant from the National Science Foundation. This financial support is gratefully acknowledged.

Uri Haber-Schaim
July 1975

CONTENTS

Chapter 1 PHYSICAL NUMBERS 1

Mathematical and Physical Numbers; Uncertainty ... Significant Digits ... Addition and Subtraction of Physical Numbers ... Computations with Physical Numbers

Chapter 2 COMPARING NUMBERS AND SETS OF NUMBERS 14

Comparing Numbers by Ordering and by Difference ... Comparing Numbers by Ratio ... The Fractional Difference: Per Cent ... Specific Quantities ... Comparing Sets of Numbers; Central Tendency ... Histograms and Frequency Distributions ... Measure of the Spread of Data

Chapter 3 NUMERICAL CALCULATIONS 53

Large and Small-Number Calculations ... Estimation ... First-Order Approximations ... An Extension of First-Order Approximations ... Relative Uncertainties in Multiplication and Division ... Finding Square Roots by an Iterative Process ... The First Order Approximation for $\sqrt{1 + \epsilon}$

Chapter 4 SLIDE RULES 67

Multiplication and Division of Powers of Two ... Non-Integral Powers of Two ... A Power-of-Ten Slide Rule ... Division and Multiplication Using Only a One-Decade Slide Rule ... Commercial Slide-Rule Scales ... Multiple Multiplication and Division ... Constant Factors, Ratios, and Uncertainty

Chapter 5 AUTOMATIC COMPUTATION 85

Programs ... Loops and Branches; Flow Charts ... Basic BASIC ... Running BASIC Programs

Chapter 6 GRAPHS 115

Functions; Independent and Dependent Variables ... Choosing Scales for Axes ... Smooth Curves and Uncertainty ... Interpolation and Extrapolation

Chapter 7 LINEAR AND POWER FUNCTIONS 139

Notation ... Homomorphic Curves ... Direct Proportions ... The Linear Function ... Finding the Equation of a Straight Line ... The Quadratic Function ... Recognizing Quadratic Functions from Graphs ... Inverse Proportions ... The Inverse-Square Function ... Recognizing Hyperbolas and Inverse-Square Functions

Chapter 8 DERIVATIVES AND ANTIDERIVATIVES 179

Function and Slope ... The Slope Function ... The Delta Process ... The Derivative of $\frac{1}{x}$ and $\sqrt{x}$ ... Some Properties of Derivatives ... Antiderivatives ... The Constant of Integration: The Initial Condition ... Short-Range Predictions ... A General Way of Calculating Integrals ... The Area Under a Curve ... The Area Function ... Numerical Integration

Chapter 9 THE EXPONENTIAL AND LOGARITHMIC FUNCTIONS 217

The Exponential Function $y = 10^x$ ... The Exponential Functions $y = b^x$ and $y = kb^x$ ... Inverse Functions ... $g(x) = \log_{10} x$: The Inverse Function of $f(x) = 10^x$ ... The Functions e^x and $\ln x$ ... The Derivative of $y = e^x$; Exponential Growth and Decay ... Recognizing Functions of the Form $y = Cb^x$ ... Recognizing Functions of the Form $y = m \log x + b$ ... Recognizing Functions of the Form $y = ax^n$ ... Scale Stretching by Logarithmic Plotting

Chapter 10 THE SINE AND COSINE FUNCTIONS 267

Sectors and Radians ... Definitions ... Symmetries and Antisymmetries ... Periodicity ... Homomorphic Trigonometric Functions ... The Functions $y = A \sin(kx)$ and $y = \cos(kx)$ ... $y = A \sin k(x - x_0)$ and $y = A \cos k(x - x_0)$ ... Recognizing Trigonometric Functions from a Graph ... Qualitative Observations on the Derivatives of $\sin x$ and $\cos x$ ... The Addition Formula for the Sine Function ... The Derivative of $\sin x$ ... The Derivative of $\cos x$

APPENDICES

Appendix 1 ALGEBRAIC MANIPULATIONS 305

Some Properties of Numbers ... Addition and Subtraction of Algebraic Expressions ... Multiplication of Algebraic Expressions ... Some Special Products ... Factoring ... Division of Polynomials in One Variable ... Algebraic Fractions ... Multiplication and Division of Algebraic Fractions

Appendix 2 WHAT CAN WE DO TO EQUATIONS? 324

Adding a Well-Chosen Zero ... Multiplying by a Well-Chosen One ... What Can Be Done to an Equation by Working With Both Sides? ... Solving the Quadratic Equation ... Substitution ... What Can Be Done to Two or More Equations? ... Graphical Solution of Two Equations ... Some Precautions

Appendix 3 INEQUALITIES 346

Notation ... Properties of Inequalities

Appendix 4 TABLES 349

Logarithms of Numbers ... Trigonometric Functions in Degrees ... Trigonometric Functions in Radians

Chapter 1. PHYSICAL NUMBERS

1.1 Mathematical and Physical Numbers; Uncertainty

Numbers mean different things in different contexts. In mathematics a number is ordinarily considered to be exact. If we refer to the number 2, we usually mean exactly 2, neither 1.99 nor 2.01, but 2.000..... carried to as many zeros as you wish to put down. Similarly, in mathematics 3.17 means 3.17000..... . To put it differently, a number in mathematics is represented by a point on the number line.

The situation is quite different when it comes to numbers which are the result of measurements. Most measurements are inexact to some extent. How inexact depends on the type and quality of the measuring instrument, and on the skill of the experimenter. The handling of such inexact numbers is a special concern of applied mathematics.

Generally, quantities such as mass, length, time, temperature, etc., are found with some sort of measuring instrument. The numerical answer is read on a scale. As a very simple example consider the measurement of the width of a piece of paper with a ruler marked in tenths of a centimeter, as shown in Fig. 1.1.

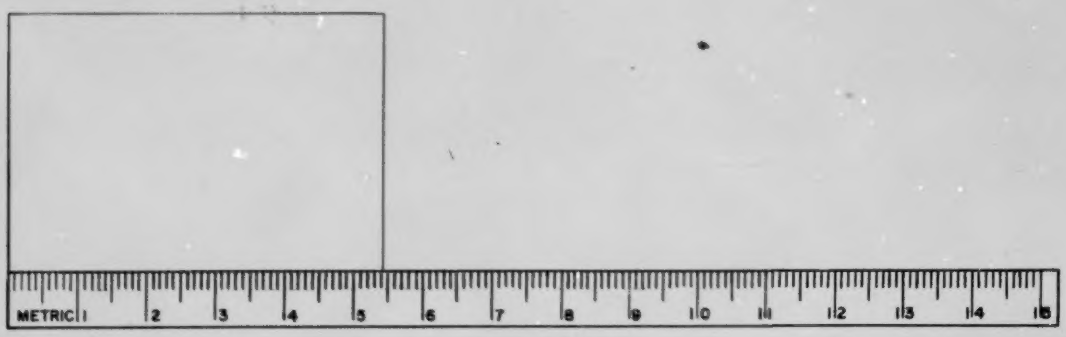


Fig. 1.1

If you showed the drawing to several people and asked them to read the ruler as carefully as possible, you would probably get a variety of answers clustered around 5.43 cm. A list of answers might be

5.41	5.44
5.43	5.47
5.42	

The last entry is obviously wrong, because the piece of paper clearly does not extend even as far as the middle of the interval between 5.4 and 5.5 on the ruler. It would be hard to argue convincingly that any one of the other answers is right and all the others wrong. The reason is that since the ruler can be read to no closer than about 0.02 cm, none of the answers are clearly incorrect except the last. It is most likely that the true value of the width of the paper lies close to the middle of the interval between 5.41 and 5.45. Expressing it differently, we can say that from the measurements the width $\underline{x}$ of the paper lies in the interval

$$5.41 < x < 5.45$$

The usual shorthand for this is

$$x = 5.43 \pm 0.02$$

When we state $x = 5.43 \pm 0.02$, we do not mean that 5.43 is the "true value" for the width of the piece of paper. All we mean is that the true value is somewhere in that interval. The interval half-width 0.02 is called the uncertainty in the number. Notice that it has a reasonable value — about how closely the ruler can be read. A number like this, which has an uncertainty resulting from measurement, is called a physical number. A physical number corresponds to an interval on the number line, and not to a point as does a mathematical number (Fig. 1.2).

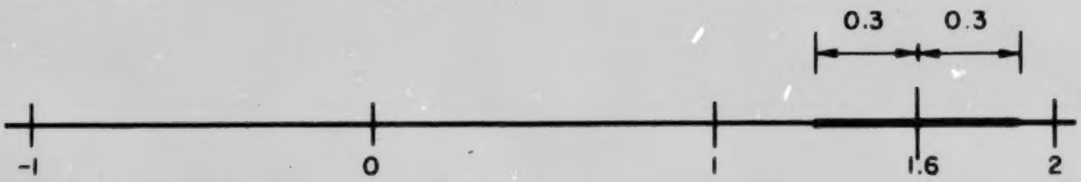


Fig. 1.2 The physical number 1.6 ±0.3 is represented by an interval on the number line. It is shown in this figure by the heavy section of the number line between 1 and 2.

Notice also that the uncertainty 0.02 is only a crude estimate, not a precise figure. It probably slightly overestimates the error, as we would wish to do in careful work. Thus the ends of the interval $5.41 < x < 5.45$ are actually somewhat "fuzzy" and we are pretty sure that the true value of $\underline{x}$ does not lie exactly at either end of the interval.

It would be nonsense, in this example, to claim an uncertainty of 0.018, or 0.023. We have no basis for claiming that much precision. We can, however, see that 0.02 is adequate while 0.01 may not be, and therefore state the uncertainty as 0.02 cm.

Some physical numbers are the result not of a single measurement, but of much scientific work. Examples are the speed of light, $(2.997925 \pm 0.000002) \times 10^{10}$ cm/sec, and the mass of an electron, $(9.1090 \pm 0.0002) \times 10^{-28}$ g. Much effort has gone into obtaining such accurate values — i.e., making the uncertainties this small.

Questions

1. Figure 1.3 shows an ammeter scale.

   (a) Read it as precisely as you can.

   (b) List your reading together with those of all your classmates. Are any of the readings obviously wrong?

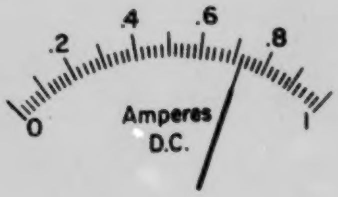


Fig. 1.3

   (c) Decide on a physical number which plausibly represents the aggregate of readings, expressing it both as an interval and in "±" form.

   (d) Compare your answers to (c) with others.

2. There are really two sources of error in reading ammeters: the reading error, as discussed in Question 1, and the inaccuracy inherent in the instrument. The latter is called a systematic error. Typically the manufacturer might certify the accuracy of an ammeter as 2 per cent of full-scale value. Taking this into consideration, what is the current measurement shown in Fig. 1.3? Express this physical number in both interval and "±" form.

3. Figure 1.4 shows the angular position of a pointer. Repeat the steps of Question 1 for this pointer.

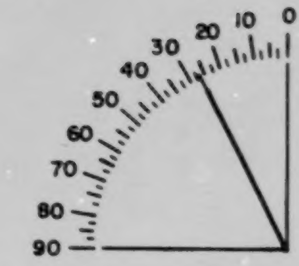


Fig. 1.4

4. Write the following physical numbers in "±" form.

   (a) 4.4 to 4.6 (c) 4.432 to 4.451

   (b) -2.1 to -2.0 (d) -1 to +7

5. Draw the section of the number line between 1 and 7.

   (a) indicate the following physical numbers on it:

   3.9 ±0.2, 3.0 ±0[illegible] 5.0 ±0.3, 2.9 ±0.1, 3.1 ±0.2

   (b) Which of the physical numbers above could possibly result from the measurement of the same object?

6. Are the "one" and "60" in the statement "one hour equals 60 minutes" mathematical or physical numbers?

1.2 Significant Digits

Writing a physical number with its uncertainty is good practice, but is sometimes cumbersome and unnecessary. For example, it may be enough for us to know that a physical number is 35 without being concerned whether the uncertainty is ±1 or ±2 or even ±3. It is general practice in such cases to state the number simply as 35, with the implied understanding that the last digit may be off either way by at most a few units. If the uncertainty happens to be ±0.1 or ±0.3, we can convey the approximate uncertainty without spelling it out, by expressing the number as 35.0. The fact that we have added another digit implies that the uncertainty is definitely less than ±1, but probably more than ±0.1. To take another example, the physical number 35.04 indicates that the uncertainty is less than ±0.1 but more than ±0.01.

Meaningless digits must be omitted in representing physical numbers in this way. The physical number 21.34 ±0.25 has an uncertainty in the third digit. Writing this physical number as 21.34 is deceptive, because this implies only an uncertainty in the fourth digit — an accuracy about ten times as great as the number actually has. This physical number should be written as 21.3. Similarly, writing 35.0 is deceptive if the uncertainty is as large as ±1.

If a physical number is written correctly, all digits are significant, except for those zeros which serve only as place-holders for the decimal point. For example, the physical number 35.18 has four significant digits; 35.0 has three (provided the number is correctly written and the zero really means plus-or-minus a few 0.1's), 35.00 has four, 0.0018 has two, the zeros to the left serving only as place-holders. Note that each of the numbers 2.4 cm, 0.024 m (meters), and 0.000024 km (kilometers), has two significant digits. This last trio demonstrates why place-holder zeros are not counted as significant.

A problem arises with numbers like 10,500 kilometers. Are the last two zeros significant digits, or are they just place-holders? For example, 10,500 kilometers is the distance from Quito, Ecuador, to Brazzaville, Congo. There is no certain way of knowing from this added information if the last two digits are significant, although the fact that they are both zero makes one suspect that they are not. It happens in this case that they are indeed not significant, since the distance was found by making measurements on a map so small that the distance could not be measured to better than about ±100 kilometers.

It is good practice to use powers-of-ten notation to show the number of significant figures of such numbers. Since $10^1 = 10$, $10^2 = 100$, $10^3 = 1000$, etc. we can write the number 10,500 ±100 as 1.05×10^4 or 10.5×10^3, or 105×10^2. In each of the three representations we see that there are three significant digits; and the power-of-ten acts as a decimal locater. Thus it would be good practice to write the Quito - Brazzaville distance in one of these forms, to indicate clearly that it is known to only about 100 kilometers accuracy.

Questions

1. Given the following physical numbers, express them in significant-digits form.
   - (a) 17.3 ±0.1
   - (b) 17.3 ±0.001
   - (c) 17.3 ±0.5
   - (d) 17.3 ±5
   - (e) 16.6 ±1
   - (f) 16.6 ±0.2
2. To how many significant digits is each of the following numbers given?
   - (a) 67.03
   - (b) 145.00
   - (c) 241.75
   - (d) 0.03001
   - (e) 4.700
   - (f) 2.75×10^5
   - (g) 2.750×10^5
   - (h) 5000
3. Given the following physical numbers in significant-digits form, give plausible equivalents in "±" form.
   - (a) 8.3
   - (b) 0.00083
   - (c) 0.0008300
   - (d) 830
   - (e) 830.0
   - (f) 2.4 cm
   - (g) 0.024 m
   - (h) 2.75×10^5 m
4. Use a centimeter scale to measure the long dimension of this page in your book and give the result in
   - (a) "±" form.
   - (b) significant-digits form.
5. In each of the following, a physical number is given without any indication of its uncertainty. Very roughly, what would you guess the uncertainty to be?
   - (a) Boston, 7 miles (highway sign).
   - (b) Centerville, pop. 1271 (sign obviously several years old).
   - (c) Yesterday's baseball attendance 10,372 (newspaper article).
   - (d) 20,000 attend mass rally (newspaper headline).
   - (e) 7.4 inches of rainfall in recent storm (weather bureau report).
   - (f) 450 calories per serving of apple pie (from an article on dieting).
   - (g) 2 pounds of coffee (from a grocery store).
   - (h) 39.37 inches in a meter (from a handbook).

1.3 Addition and Subtraction of Physical Numbers

The addition of the two mathematical numbers 7.9 and 5.6 obviously gives 13.5. Now consider the addition of 7.9 ±0.2 grams of salt to 5.6 ±0.1 grams of salt. The result could be as large as $(7.9 + 0.2) + (5.6 + 0.1) = (7.9 + 5.6) + (0.2 + 0.1) = 13.5 + 0.3$ g, or it could be as small as $(7.9 - 0.2) + (5.6 - 0.1) = (7.9 + 5.6) - (0.2 + 0.1) = 13.5 - 0.3$ g. The result is thus 13.5 ±0.3 g; we note that the uncertainties have added.

By considering in this fashion the largest and smallest values the result could have, we find the corresponding general rule:

$$(A \pm a) + (B \pm b) = (A + B) \pm (a + b) \qquad (1)$$

Similarly, if 5.6 ±0.1 g of salt is taken away from 7.9 ±0.2 g, the amount remaining could be as large as $(7.9 + 0.2) - (5.6 - 0.1) = (7.9 - 5.6) + (0.2 + 0.1) = 2.3 + 0.3$ g, and could be as small as $(7.9 - 0.2) - (5.6 + 0.1) = (7.9 - 5.6) - (0.2 + 0.1) = 2.3 - 0.3$ g. The difference of these two physical numbers is thus 2.3 ±0.3 g, and we see that uncertainties add in subtraction as well as in addition. (Since subtraction is equivalent to addition of the negative, we could have deduced this from our earlier formula for addition.) We have thus

$$(A \pm a) - (B \pm b) = (A - B) \pm (a + b) \qquad (2)$$

This completes the formulation of the rules for addition and subtraction of physical numbers. However, in practice there are special cases worth considering. Suppose first that one of the uncertainties is much smaller than the other — say $\underline{b}$ is much smaller than $\underline{a}$. This may be written $b \ll a$. (Note that by convention $\underline{a}$ and $\underline{b}$ are both positive.) Then the uncertainty may be taken simply as $\underline{a}$. For example, consider the sum (5.3 ±0.2) + (3.418 ±0.003). It would be rather silly to write the result as 8.718 ±0.203, since the first number is known only to within 0.2, an additional uncertainty of 0.003 is meaningless. We would ordinarily write the uncertainty as simply ±0.2. If the result is written in significant-digits form it should be written 8.7, not 8.718 nor even 8.72. Note two things that have happened: the larger uncertainty has "swamped" the smaller, and some significant digits in the more accurate number have lost their significance in the sum.

In the example we have been discussing, to illustrate the $b << a$ situation, A and B had comparable magnitudes. More usually when $b << a$ we have also $|B| << |A|$. Consider for example (5.31 ±0.03) + (0.0128 ±0.0001). The result is 5.32 ±0.03, and one must accept the necessity of throwing away the last two digits of B, which have become insignificant in the sum. Much as one might wish to write the result as 5.3228, this would be quite misleading.

In extreme cases B can be totally "swamped" by the uncertainty in A. Consider for example the addition of the physical numbers 3.7 ±0.1 and 0.016 ±0.002. This might arise in the following way: The thickness of a steel plate is measured with a ruler and found to be 3.7 ±0.1 mm. The thickness of aluminum foil is found with a micrometer to be 0.016 ±0.002 mm. Then the plate and the foil are pressed together, and the combined thickness is measured with a ruler. The result of this final measurement would probably be 3.7 ±0.1, the same as the first measurement.

The point is that the uncertainty in the steel plate is already about six times the thickness of the aluminum foil. Thus, adding the foil to the plate does not measurably (using a ruler) increase its thickness or the uncertainty of the measurement.

Next, irrespective of the relative sizes of the uncertainties, let us consider the effects of the relative sizes of A and B. If these are of nearly the same size, nothing remarkable happens when they are added; however a dramatic loss in significant digits can occur when one is subtracted from the other. Thus 29.27 - 29.18 = 0.09 is a calculation in which three significant digits are lost.

In general terms, in the subtraction $(A \pm a) - (B \pm b) = (A - B) \pm (a + b)$ it can happen that $(A - B)$ has a magnitude much less than either A or B, and perhaps comparable with or even less than the uncertainty $(a + b)$. It is important to recognize this loss of significant digits when finding the difference of nearly equal numbers.

We may summarize our discussion in three "rules-of-thumb" for the addition and subtraction of physical numbers:

1. When uncertainties differ widely, the larger one governs.

2. Don't save digits which have become insignificant.

3. Be on guard to detect the loss in significant digits which occurs when taking the difference of nearly equal numbers, and if possible avoid the necessity of such a calculation.

Questions

1. Add each of the following pairs of physical numbers.

   (a) $(2.71 \pm 0.03) + (0.01 \pm 0.01)$

   (b) $(47.8 \pm 0.1) + (1000 \pm 1)$

   (c) $(0.007 \pm 0.001) + (0.0003 \pm 0.0001)$

   (d) $(63 \pm 1) + (2 \pm 0.5)$

   (e) $(8 \pm 1) + (11 \pm 3) + (14 \pm 2)$

   (f) $(3.7 \pm 0.1) + 10 \times (0.016 \pm 0.002)$

   (Part (f) corresponds to adding ten sheets of aluminum foil to the steel plate discussed in the text.)

2. Calculate the answer to each of the following operations involving physical numbers.

   (a) $12.5 + 26.8$

   (b) $12.5 + 2.68$

   (c) $12.5 + 0.0268$

   (d) $26.8 + 12.5 + 1.32$

   (e) $12.5 + 26.8 + 1.32$

   (f) $2.5 \times 10^2 - 1.8 \times 10^3$

   (g) $1.01 \times 10^3 - 9.8 \times 10^2$

   (h) $6.31 \times 10^5 + 2.12 \times 10^2$

3. One technique for weighing an animal is to weigh oneself on a bathroom scale while holding the animal, and while not. Explain why this technique works better for a large dog than for a small kitten.

4. Pediatricians sometimes advise new parents not to weigh their baby too frequently. Can you think of a reason for this advice?

1.4 Computations with Physical Numbers

Wishing to measure the density of a fluid, you determine that a volume of 151 ±2 cm^3 has a mass of 212.1 ±0.5 g. Nominally the density is then 212.1/151 g/cm^3, but what is the uncertainty?

This brings up the difficult but important question of carrying out calculations with physical numbers beyond additions and subtractions. To draw conclusions from experiments, some calculation is usually required, and it is important to know reliably the uncertainty in the result. In the present example the most obvious approach is to calculate the smallest and largest values the result can have. The smallest possible value, obtained by making the numerator as small and the denominator as large as possible, is $211.6/153 = 1.383$ g/cm^3. By similar reasoning the largest possible value is $212.6/149 = 1.427$ g/cm^3. Thus the answer lies in an interval of length $1.427 - 1.383 = 0.044$ and the center of the interval is $\frac{1}{2}(1.427 + 1.383) = 1.405$. Thus the density would be given as 1.405 ±0.022 g/cm^3.

It is quite all right to state this result as 1.40 ±0.03, enlarging the interval slightly in order to simplify the answer, if one is not concerned with obtaining the closest possible estimate. It is, however, incorrect and misleading to state the result as simply 1.405 g/cm^3 without stating the uncertainty; this would imply four significant digits of accuracy, whereas we really have at best only three.

Sometimes this amount of care is not needed in computations; we may be satisfied with a general indication of the uncertainty of a result, rather than a strictly correct interval. In this case significant-digit form may suffice for physical numbers entering the calculation, and it may be possible to infer how many digits should be kept in the final result, so as neither to sacrifice truly significant information by quoting too few digits, nor to imply more information than is actually present by quoting too many. In general, one rule of thumb should be kept in mind: it is unusual for the number of significant digits to increase during a calculation.

When more care is required, i.e., when one wants really to know the interval in which the result of a calculation lies, perhaps the best general advice is: (1) Work with the numbers in interval form, not significant-digit form, as the latter is too crude for this purpose; (2) Calculate the smallest and the largest value the answer could have, as in the example we gave; (3) Be careful not to introduce additional errors by carrying too few places in the calculation.

This last point is known as carrying "guard digits" to prevent round-off error. Round-off error, as its name suggests, is the error incurred when a decimal is truncated or rounded, as will continually occur in a computation of any length. While it is misleading to state a result to more apparent significant digits than really known, there is absolutely nothing wrong with carrying extra digits in the intermediate stages of a calculation, and in fact this is recommended. This occurs very commonly when the computations are done by a computer.

The product of two physical numbers occurs so often that it deserves special consideration. Suppose we wish to compute $C \pm c = (A \pm a)(B \pm b)$, and assume that $\underline{A}$ and $\underline{B}$ are positive with $a < A$ and $b < B$. Then the smallest and largest values of the product are $(A - a)(B - b)$ and $(A + a)(B + b)$; $\underline{C}$ is the average of these and $2c$ the difference, so that

$$(A \pm a)(B \pm b) = C \pm c = [AB + ab] \pm [aB + bA] \tag{3}$$

Frequently the term $\underline{ab}$ in Equation (3) will be so small compared to $\underline{AB}$ that it can be neglected. Thus, the uncertainty in $\underline{AB}$ equals $(aB + bA)$; the uncertainty in $\underline{A}$ times $\underline{B}$, plus the uncertainty in $\underline{B}$ times $\underline{A}$.

We shall return to the matter of the behavior of uncertainties in multiplication and division in Chapter 3.

Questions

1. Using the measured values

$$x = 1.1 \pm 0.1$$
$$y = 0.5 \pm 0.1$$
$$z = 2.0 \pm 0.2$$

compute carefully each of the following physical numbers:

(a) $x^2 + y$ (b) $\frac{x - 2y}{z}$ (c) $\frac{2x + 3y}{x - y}$

2. A crude estimate of the mean radius of the earth is 6400 ±100 km.

(a) What is the resulting value of its volume?

(b) Given that the earth's mass is $6.0 \pm 0.1 \times 10^{27}$ g, calculate its mean density in g/cm^3.

3. The piece of paper in Fig. 1.1 was determined to have a width of 5.43 ±0.02 cm. Suppose its length is measured to be 6.44 ±0.02 cm. Assuming that its shape is perfectly rectangular, calculate its area as a physical number.

4. By plotting $\underline{A}$ and A + a horizontally, and $\underline{B}$ and B + b vertically, interpret Equation (3) graphically in terms of the areas of various rectangles.

5. How should Equation (3) and the accompanying discussion be modified if one of the factors, say $\underline{A}$, is negative?

6. Many calculations involve both mathematical and physical numbers. Suppose the radius $\underline{r}$ of a circle is 5.0 ±0.1 cm. Compute its circumference $L = 2\pi r$. Are the numbers 2 and π physical or mathematical numbers? To how many decimal places need π be expressed in this calculation?

7. If $\underline{a}$ is much smaller than the magnitude of $\underline{A}$, show that the magnitude of the uncertainty in the reciprocal of $A \pm a$ is approximately $\frac{a}{A^2}$.

8. The power, in watts, consumed by an electric circuit is the product of voltage E and current I, in volts and amperes respectively. Suppose voltage is measured by a voltmeter accurate to ±2 volts, and current by an ammeter good to ±0.03 amp. What power corresponds to each of the following pairs of nominal readings? (Be sure to give your answer as physical numbers.)

(a) $E = 110$, $I = 1.25$ (c) $E = 2$, $I = 5.51$

(b) $E = 115$, $I = 0.04$ (d) $E = 2$, $I = 0.04$

9. Use a centimeter scale to find the area of the cover of this book in

(a) cm^2 (square centimeters).

(b) mm^2 (square millimeters).

10. Suppose that the base of a certain Egyptian pyramid is found upon measurement to be very nearly a square 100 ±2 meters on a side. The height is measured to be 100 ±5 meters. A piece of stone taken from it having a mass of 357.5 ±0.2 grams is found to displace 100 ±3 cubic centimeters of water. What is the total mass of the pyramid?

Chapter 2. COMPARING NUMBERS AND SETS OF NUMBERS

2.1 Comparing Numbers by Ordering and by Difference

In Egyptian mythology the souls of the dead were weighed in a balance against an ostrich feather. For salvation it was crucial that the soul be heavier than the feather—but it didn't matter by how much. This is an example of comparison by ordering, where the only information required is which of two numbers is larger.

More down-to-earth examples where numbers are compared by simple ordering are readily found: Furniture movers need only to know which is larger, the width of a door or the width of a piano, in order to decide whether the piano can be taken by that route. In selecting a portion of food one might pick the largest if one is hungry; or the smallest if on a diet. In a track meet the broad-jump event is won by the longest jump, no matter how little this jump exceeds the second best.

Sometimes, however, just ordering numbers is not enough to tell us what we wish to know. For example, in following a baby's growth, one is probably interested not only in the fact that the baby's weight is greater than it was a year ago, but in how much greater. As another example, suppose you were deciding from which of two dealers to buy a certain automobile. If one price was substantially lower than the other (say, several hundred dollars) you would probably choose that one. But if the prices were nearly the same, the decision might be made on other grounds—for example, convenience and reliability of service. Here we clearly must know not only which is cheaper but also by how much.

To make this sort of comparison we calculate the difference between the two quantities by subtracting one from the other. Of course, since a - b does not equal b - a, we have to decide which difference to use. It is particularly important to be consistent as to which number is subtracted when we describe a change in a given quantity. Consider the change in hourly temperature readings or daily stock-market quotations. One always gives the change in going from the earlier to the later reading, and therefore subtracts the earlier reading from the later one. A change in temperature of 5°C means that the temperature increased 5°C. A change of -3 points on the stock market means in everyday language that the market dropped by three points.

Expressing the change in a quantity by the later value minus the earlier is so common that it is designated by a special symbol, the Greek capital letter Δ (delta). For example, if t represents temperature, Δt stands for the change in temperature, i.e., later temperature minus earlier. Since Δt can be positive or negative it can be used to express both increases and decreases.

If two quantities are to be compared by taking their difference, they must have the same units, or be converted into the same units. For example, there is no sense in subtracting one inch from three feet to get two as the difference in length. One can get the difference in length by converting three feet to 36 inches, and subtracting one inch from that. The resulting length difference of 35 inches does make sense.

In finding the difference between two physical numbers, it is important to keep in mind the warning in the previous chapter about possible loss of significant digits. Consider, for example, $(5.46 \pm 0.02)\text{cm} - (5.43 \pm 0.02)\text{cm}$, which might be the difference in the widths of two pieces of paper. The result, 0.03 ± 0.04 cm, is inconclusive in telling which piece is wider. Note that whereas the original values were good to three significant digits, the difference hardly has one significant digit of accuracy.

Questions

1. In which of the following situations would you be satisfied with a comparison by ordering? In which by taking a difference? In which would neither form of comparison be useful?

   (a) Deciding whether a book will fit into a certain shelf of a bookcase.

   (b) Choosing among cabinets to fit into a kitchen, making as close a fit as possible.

   (c) Deciding whether your reducing diet is going well.

   (d) Describing the height gain of a child over a one-year period.

   (e) Selecting the baseball player with the highest batting average.

   (f) Deciding which of two baseball players, with known batting averages, to hire for a team.

2. Suppose that Consumers Research measured the following gas mileage figures for six sample new cars:

   (a) 13.7 (miles per gallon)
   (b) 12.8
   (c) 14.1
   (d) 11.9
   (e) 13.9
   (f) 13.2

   Which model performs best? From these data, does it appear that gas mileage will be an important criterion in choosing which model to buy? Make up a hypothetical new set of data which would change your answer to this problem.

3. Table 2.1 gives the rate of unemployment in the United States, as the number of unemployed per 100 workers, for the years 1963 through 1971.

   (a) When was unemployment per 100 workers greatest?

   (b) When was it least?

   (c) What kind of comparison did you use in deciding on your answers to (a) and (b)?

TABLE 2.1

Year	Unemployed per 100	Change in Unemployment
1963	5.7	
1964	5.2	____
1965	4.5	____
1966	3.8	____
1967	3.8	____
1968	3.6	____
1969	3.5	____
1970	4.9	____
1971	5.9	____

4. (a) Fill in the third column of Table 2.1 with the change in the unemployment rate per 100 workers between each two consecutive years.
(b) Between which two consecutive years did unemployment increase most rapidly? Decrease most rapidly?

5. Give an example of a comparison of two nearly equal physical numbers where almost all significant digits are lost by taking the difference.

2.2 Comparing Numbers by Ratio

Often numbers are compared by stating how many times one is larger than the other, rather than by how much they differ. For example, a 100 cm rod is 50 times longer than a 2.0 cm piece of chalk, though the difference between them is 98 cm. To find how many times a is larger than b, we calculate the ratio of a to b, i.e., we divide a by b. Like subtraction, division is not commutative: $\frac{a}{b}$ does not equal $\frac{b}{a}$, and thus one must be careful which one uses. If we speak of the ratio of a to b, we mean $\frac{a}{b}$. The ratio $\frac{b}{a}$ on the other hand, is the ratio of b to a.

If two quantities are to be compared by finding their ratio, they must be expressed in the same units, as is the case when two quantities are compared by taking their difference. If they are given in different units, one of them has to be converted to the units of the other. Consider the following example: It takes one commuter 1 hour and 12 minutes to commute to work; another commuter gets to work in 27 minutes. How many times longer does it take the first person to get to work than it takes the second person? First, converting hours into minutes, 1 hour = 60 minutes, and the first man takes 60 + 12 minutes = 72 minutes. The ratio yields $\frac{72 \text{ min.}}{27 \text{ min.}} = 2.7$ times longer.

In the last example it would also make sense to compare the times it takes the commuters to get to work in terms of difference: It takes the first 72 minutes - 27 minutes = 45 minutes longer. However, if it takes an airplane 20 minutes and a cyclist needs 40 ±2 hours to cover a given distance, the difference in times would be about 40 hours. This would also be true if it took the airplane 30 minutes. A comparison by ratio shows $\frac{40 \times 60 \text{ min.}}{20 \text{ min.}} = 120$ in one case and $\frac{40 \times 60 \text{ min.}}{30 \text{ min.}} = 80$ in the other, a significant difference; here the ratio carries information which the difference does not.

As in comparing physical numbers by difference, when comparing them by finding their ratio we must pay attention to significant digits. For example, the ratio of the lengths of two nails, one 5.52 cm long and the other 2.3 cm, is $\frac{5.52 \text{ cm}}{2.3 \text{ cm}} = 2.4$, a physical number having only two significant digits.

The idea of order of magnitude is essentially related to comparing numbers by ratio. It is particularly useful in discussing very large or very small quantities. If the ratio $\frac{a}{b}$ of two positive numbers a and b is about 1 (say, between $\frac{1}{2}$ and 2), we say that a and b are of the same order of magnitude. If a is about 10 times b (or between 5 and 20 times), it is said to be one order of magnitude larger than b. If the ratio is about $100 = 10^2$, the numbers differ by two orders of magnitude.

To illustrate the usefulness of orders of magnitude, we show in Table 2.2 the masses of the sun and some of its planets measured in units of earth mass.

TABLE 2.2

Sun	3.3×10^5	Mars	1.1×10^{-1}
Mercury	5.5×10^{-2}	Jupiter	3.1×10^2
Venus	0.81	Saturn	0.94×10^2
Earth	1.00	Uranus	1.4×10
Moon	1.2×10^{-2}	Neptune	1.7×10

Clearly, the masses of Earth and Venus are of the same order of magnitude, that of Mars being one order of magnitude lower and Saturn two orders of magnitude higher.

We see from Table 2.2 that Jupiter's mass is between two and three orders of magnitude larger than Earth's. Here we are in a gray area; the two do not differ by two orders of magnitude, nor do they differ by three orders. The idea of order of magnitude is thus highly approximate.

Frequently this lack of precision, or "fuzziness" in the idea of order of magnitude, is not at all a disadvantage. On the contrary, it can be just what we need to express a value which is fuzzy by nature. Consider, for example, the question of how long man, *homo sapiens* that is, has existed on earth. Anthropologists and archeologists differ in their interpretation of the very fragmentary evidence which has been found, and moreover (at least according to the bulk of scientific opinion) the evolution of man was probably a gradual process, in which no precise transition point can be convincingly demonstrated. To say that man's tenure on earth has been of the order of magnitude of one million years expresses our state of knowledge of this value well; the values two million and 500,000 years are not ruled out, as in fact they should not be.

Questions

1. If $\underline{a}$ is twice as big as $\underline{b}$, what can you say about the ratio of $\underline{b}$ to $\underline{a}$? About the ratio of $\underline{a}$ to $\underline{b}$?
2. If $\underline{a}$ is bigger than $\underline{b}$, and $\underline{c}$ is bigger than $\underline{d}$, what can you say about the ratio of $\underline{a}$ to $\underline{b}$, as compared to the ratio of $\underline{c}$ to $\underline{d}$?

3. If a is bigger than b and b is bigger than c, and all these numbers are positive, what can you say about the ratio of a to b, as compared to the ratio of a to c? What can you say if all the numbers are negative?

4. The ages of two brothers have a constant difference. What happens to the ratio of their ages as they grow older?

5. In which of the following situations do ratios provide the best form of comparison? In which would taking the difference between the two quantities be more meaningful? In which would you merely use ordering?

   (a) The sizes of two armies engaged in battle.

   (b) The weights of two opposing football linemen.

   (c) The sizes of two families.

   (d) The altitude of an airplane and the height of a mountain over which it is about to fly.

6. Brand A beer claims to have 21 million bubbles in a bottle, to Brand X's 20 million. Compared by difference, this is a million bubbles more for Brand A; compared by ratio, Brand A has 1.05 times as many bubbles as Brand X. Which comparison do you think Brand A will put into its advertising (assuming that bubbles are a good thing)? Why? Which is the most meaningful mode of comparison in this case?

7. Brand B cigarettes claim to have two micrograms of tar and nicotine in each cigarette to Brand Y's three micrograms. (A microgram is a millionth of a gram.) Give the comparisons by difference and ratio, and answer the same questions as for Brand A beer in the preceding question.

8. Advertising and public relations provide many examples of poor modes of comparison. Why? Find two or three examples of numerical comparisons from these sources, explain why they were done the way they are, and argue for their relevance or irrelevance.

9. A large meteorite has a mass of 5×10^4 kilograms. The earth has a mass of 5.983×10^{21} metric tons. (One metric ton is 1,000 kilograms.) What is the ratio of the mass of this meteorite to the mass of the earth? How many orders of magnitude larger is the earth's mass than the mass of the meteorite?

10. Suppose that in the two quadratics

$$P = a_1 + a_2x + a_3x^2$$
$$Q = b_1 + b_2x + b_3x^2$$

the coefficients a_1, a_2, a_3, b_1, b_2, and b_3 all have order of magnitude 1.

(a) If $\underline{x}$ has order of magnitude 10^{-8}, what are the orders of magnitude of P, Q, PQ, and $\frac{P}{Q}$? Write approximations for these four quantities.

(b) Answer the question of part (a) if $\underline{x}$ has order of magnitude 10^8.

2.3 The Fractional Difference: Per Cent

Suppose that in six months a baby's weight increased from 15 lb to 25 lb and the weight of a boy increased from 60 lb to 70 lb. In both cases there was a change in weight of 10 lb, yet from a practical point of view the two changes are quite different; for the baby it means new clothes, but for the boy it probably does not. This is true because the change in weight of the baby is a much larger fraction of its original weight, $\frac{25 - 15}{15} = 0.67$, whereas in the case of the boy $\frac{70 - 60}{60} = 0.17$. This method of comparison has something in common with both the preceding methods. The numerator $\Delta w = 25 - 15$ is the difference between the baby's earlier and later weights, i.e., his change in weight; the denominator, 15, is the weight he started out with. The entire expression $\frac{\Delta w}{w} = \frac{25 - 15}{15}$ is the ratio of the change in weight to the original weight.

The quantity $\frac{b - a}{a}$ is called the fractional or relative difference. It provides a means of judging whether a difference is large or small, compared with an original or base quantity.

Since a and b must have the same units to make the subtraction meaningful, the quotient $\frac{b - a}{a}$ is a pure number independent of the units in which a and b were expressed. Note also that it must be clear whether you mean $\frac{b - a}{a}$ or $\frac{b - a}{b}$, i.e., whether you are comparing the difference with a or with b. With changes in the same quantity, as with the growing baby, it is the earlier or original value against which the comparison is made. "The fractional difference by which 90.0 differs from 80.0" means $\frac{90.0 - 80.0}{80.0} = 0.125$, and "the fractional difference by which 80.0 differs from 90.0" is $\frac{80.0 - 90.0}{90.0} = -0.111$. These two fractional differences are of course not equal; thus one must be careful to avoid ambiguity in dealing with fractional differences.

Obviously, the denominator cannot be zero; for example, it is meaningless to talk about the fractional increase in profits during the first year of operation of a new business.

Another way of looking at a fractional difference is that $\frac{b - a}{a}$ expresses the difference per unit of a, i.e., how much the quantity changes for each unit of it that was there originally. For example, for each pound of baby that you started out with, you ended up with 0.67 pounds extra at the end. As we shall see later the word per is generally associated with divisions.

Often it is useful to express a fractional difference not per unit but per hundred units; in fact, this is the usual practice. The result is referred to as the "percentage difference." In the case of the growing baby, the fractional difference, 0.67, in its two weights corresponds to $0.67 \times 100 = 67$ per cent.

Percentage is frequently used to express concentration. Thus a nut mixture containing 20 per cent cashews has 20 pounds of cashews in each 100 pounds. A 5 per cent salt solution is usually defined to be one containing 5 g of salt in each 100 g of solution. We could equally well say 5 pounds of salt in each 100 pounds, or simply 5 units of salt in each 100 units, or 5 units per 100. (The term "per cent," in fact, comes from the Latin for "for each 100.")

When we use a percentage to express a fractional difference, we are stating the number of units of the difference corresponding to each 100 units of the original quantity. A weight gain of 5 per cent is a gain of 5 units for each 100 units of the original amount (or 0.05 units for each 1 unit). If the original amount was 50 pounds, then a 5 per cent gain would amount to (0.05) × 50 = 2.5 pounds. If the original amount was 300 grams, then a 5 per cent gain would be a gain of (0.05) × 300 = 15 grams.

Percentage differences are frequently used to express the uncertainty of a physical number. For example, 50 ±3% means that the uncertainty is 3 per cent of 50, or (0.03) × 50 = 1.5. Thus 50 ±3% = 50 ±1.5. These two forms expressing uncertainties are called relative uncertainty (expressed here in per cent) and absolute uncertainty respectively.

To convert from relative to absolute uncertainty one carries out the steps

$$A \pm p\% = A + (0.01)\, pA$$

Conversion from absolute to relative uncertainty is given by

$$A \pm a = A\left(1 \pm \frac{a}{A}\right)$$

For example

$$50 \pm 1.5 = 50\left(1 \pm \frac{1.5}{50}\right) = 50(1 \pm 0.03) = 50 \pm 3\%$$

Questions

1. A 12-pound baby eats a four-ounce jar of baby food for a meal. His 160-pound father eats a total of one pound of food for a meal.
    (a) How much does each eat relative to his body weight?
    (b) Which eats more relative to his weight than the other?
2. Between the years 1950 and 1960, the population of Arizona increased from 750,000 to 1,302,000. In the same period of time, the population of Arkansas went from 1,910,000 to 1,786,000.
    (a) What was the change in the population of each state?
    (b) What was the ratio of increase to initial population?
    (c) What was the increase in population per 1000 people?
    (d) What was the relative change in population?

3. In a given year A receives $300 in interest on $6000 in a savings account. In the same year, B receives $200 on $3500 in a savings account in a different bank. Which bank pays the higher rate of interest?

4. A's salary was raised from $10,000 to $11,000 per year, and B's salary from $15,000 to $16,300 per year. How would you compare their raises?

5. By what per cent does 90.0 differ from 80.0? By what per cent does 80.0 differ from 90.0? Answer the same questions for 100.0 and 200.0.

6. By what fractional or per cent difference does 1.00 meter exceed 1.00 yard? What is their ratio? (1 inch = 2.54 centimers, exactly; this is the definition of the inch.)

7. Suppose you read that a newspaper's circulation increased by 5,025 in one year.

   (a) Does this figure tell you that the newspaper's circulation increased significantly during the year?

   (b) How would you answer part (a) if you knew that the circulation at the end of the year was 20,100? Was 2,010,500?

8. Some numbers and their relative uncertainties are given below. How many digits are significant in each of the numbers?

   (a) 1.37492 to 1%
   (b) 2.30476 to 0.02%
   (c) 2.3 to 0.02%
   (d) 0.005982 to 5%
   (e) 100.1 to 1%

9. In the preceding problem, express each of the numbers using absolute uncertainties. Omit meaningless digits.

10. Express each of the following numbers using per cent uncertainties:

   (a) 100 ±3
   (b) 250 ±5
   (c) 250 ±1
   (d) 200 ±100
   (e) -0.5 ±0.05

11. Fly-by-Night Airlines announces a 2000 per cent increase in passenger miles flown this year over last year. What other information would be required for a meaningful assessment of the situation?

12. Return to the problems on beer and cigarettes at the end of the preceding section. Calculate the fractional differences. Is this a meaningful mode of comparison in either case?

13. Amalgamated Goosefeathers sold 10,000.0 bushels of the product this year, a 50 per cent increase over last year. How many did they sell last year? If their sales were a 50 per cent decrease, how many did they sell last year?

14. The first steel mill in a new country was built this year, and has produced five tons. What is the most meaningful way of comparing this year's steel production with last year's? What problem arises with comparison by ratio and by fractional difference?

15. A is 100.0. B is 10.0 per cent larger than A. C is 10.0 per cent larger than B. How much larger is C than A?

16. A merchant sells a certain item at a retail price 50 per cent greater than the wholesale cost. During a sale the retail price is reduced by 20 per cent. What percentage profit does the merchant make on that item during the sale?

17. (a) For any two numbers A and B, find the general relation between their ratio and their fractional difference. Express the relation in words.

    (b) Find the general relation between their difference and their fractional difference and express it in words.

2.4 Specific Quantities

If you are told that a large can of a certain brand of peas costs 37 cents and a small can costs 19 cents, you cannot judge which is the better buy. You need to know the amount of peas in each can. If you find out that the 37-cent can contains 17 ounces and the 19-cent can contains 8.5 ounces, you are in a good position to choose between them. You divide the price by the weight to get the cost per ounce for each can. For one can this is $\frac{\text{37 cents}}{\text{17 ounces}}$ = 2.18 cents per ounce (often written 2.18 cents/ounce) and for the other it is $\frac{\text{19 cents}}{\text{8.5 ounces}}$ = 2.24 cents per ounce (2.24 cents/ounce). These two numbers represent the cost of one ounce of peas and can be compared to find out which is the better buy. In this example the cost of one ounce of the contents of the large can is less than the cost of one ounce of the contents of the small can, so the large can is the better buy. The cost of peas per ounce is called a specific quantity. (We are assuming, of course, that both cans contain the same brand and quality of peas, and that you can use all the peas in the large can.)

In calculating the cost per ounce for peas we divide the cost of a can of peas by the weight of the peas in the can. The fact that we divide one number by another does not mean that we have taken a ratio. In fact we have not. We have a ratio only when we divide two numbers that have the same units. When we find the cost per ounce of something, the two numbers we divide are given in different units and the result we get is meaningless unless we state the units with the number. To say, "The price of peas is 2.24" is nonsense. To say, "The price of peas is 2.24 cents" does not make sense either. To say, "The price of peas is 2.24 cents per ounce" makes sense. Once we know the price of peas from the two cans in the same units, namely cents per ounce, we can compare the two prices by any of the methods of comparison we have discussed.

Questions

1. A 15-ounce (net contents) can of peaches costs 23 cents, while a 29-ounce can costs 35 cents. Compare the two costs in terms of cents per ounce. Which is better?

2. (a) Five pounds of salt are dissolved in three gallons of water. How many pounds of salt per gallon of water are in the resulting solution?

   (b) Seven pounds of salt are added to five gallons of water. Is this solution saltier than that in (a)? How much saltier?

3. In 1967 it was estimated that in metropolitan areas (cities of 250,000 or more) there were 2,631,000 poor whites and 1,833,000 poor nonwhites out of a population of 23,824,000 whites and 3,184,000 nonwhites. What informative comparisons can you make using these four quantities?

4. A group of 50 people is in a room of 6.0 m × 8.0 m × 2.5 m. Another group of 60 people is in a room of 7.0 m × 7.0 m × 4.0 m. In which room are the people more crowded?

5. In the text we divided the cost of the can by the weight of peas. It would have been possible to do it the other way around, and get (for the larger can), the specific quantity $\frac{17 \text{ ounces}}{37 \text{ cents}} = 0.46$ ounces for each cent, or 0.46 ounces per cent, or 0.46 ounces/cent. Does this quantity mean anything, and if so what? Can you think of any advantages it might have over the quantity calculated in the text?

6. Compare the two cans of peaches in Question 1 in terms of ounces/cent.

7. In the text the quantities 37 cents and 17 ounces are used. Are these mathematical or physical numbers?

8. (a) Two hundred and fifty marbles have a total mass of 2750 g. What is the mass per marble (the average mass of one marble)?

   (b) Another collection of 150 marbles has a mass of 1800 g. What is the average mass per marble in this case?

   (c) In which collection of marbles is the mass per marble greater? How many times greater?

9. Many things cost more in smaller quantities: For example, a coal company charges $35 for a half-ton of coal and $60 per ton for quantities of a ton or more. What is the specific cost in the two cases? What possible reasons are there for this practice of higher unit costs for small quantities?

10. Sometimes, instead of expressing things per unit, or per hundred units it is useful to express them per million units. This is especially true in biological applications, where quite dilute, minute quantities of some substances can have substantial effects. The term used is "parts per million" (ppm). Given a 5 per cent salt solution, express its concentration in ppm. Given a 15 ppm solution of vitamin B, express its concentration in per cent. What is the general relationship between per cent and ppm?

2.5 Comparing Sets of Numbers; Central Tendency

In a lifetime test, one light bulb of Brand A lasted for 1242 hours, and one of Brand B lasted 1073 hours. What can we conclude from this?

The fractional difference of the Brand A over the Brand B sample is $\frac{(1242 - 1073)}{1073} = 0.158$. That is to say, the A sample lasted about 16 per cent longer than the B sample. However, on the basis of only this pair of samples, we can say virtually nothing about the relative performance of Brand A and Brand B in general.

Suppose then, to attempt to answer this more general question, a lifetime test is carried out on a sample of 25 bulbs of each brand, with the results shown in Table 2.3:

TABLE 2.3. BULB LIFETIMES IN HOURS, RAW DATA

Brand A		Brand B	
1242	893	1073	1041
1013	1167	1304	1251
869	998	1243	1462
1149	1417	1471	1653
973	1091	1169	1204
1160	1009	941	772
844	897	1368	1309
1302	1026	1265	1261
1033	1140	1141	1575
1125	839	1322	1381
741	1026	1278	1320
1087	940	1404	1135
1003		1289	

By studying these results we can begin to see a trend of longer lifetimes in the B samples, but the situation is far from clear. We wish to discuss how data such as these can be organized and presented in order to bring out more clearly whatever information is present, and how such data can be characterized or summarized in brief forms more suited for comparison.

The first thing which might occur to one is to compute the average or mean of each set of 25 numbers. This of course is simply the sum of the numbers divided by 25. This calculation is rather tedious, unless one has a desk computer handy.

Brand A mean:	1039 hours
Brand B mean:	1266 hours

Now, contrary to our first impression stemming from a comparison of only one pair of light bulbs, it looks as though Brand B may be superior.

In order to be able to describe mathematically such operations as computing averages, we introduce the notation a_1, a_2, a_3, ... a_{25} for the Brand A values in Table 2.3, and b_1, b_2, b_3, ... b_{25} for the Brand B values. Thus a_1 = 1242 hours, a_2 = 1013 hours, ... b_{25} = 1135 hours.

Then the average or mean $\overline{a}$ of the Brand A group is

$$\overline{a} = \frac{a_1 + a_2 + a_3 + \ldots + a_{25}}{25}$$

This may be written in summation notation as

$$\overline{a} = \frac{1}{25} \sum_{k=1}^{25} a_k$$

The symbol $\sum$ is a capital sigma, the Greek S, standing for summation. The k is called the index of summation, and here k is said to "run" from the lower limit 1 to the upper limit 25. The value of $\sum_{k=1}^{25} a_k$ is obtained by taking the value of a_k for k = 1, adding the value of a_k for k = 2, adding the value of a_k for k = 3, etc. until k = 25 has been reached.

The general formula for the mean of N numbers $(x_1, x_2, x_3, \ldots x_N)$ in this notation is:

$$\overline{x} = \frac{1}{N} \sum_{k=1}^{N} x_k \qquad (1)$$

Let us continue our quest to visualize the data better. As a first step, we might re-list each set of numbers in order of increasing value. The result is shown in Table 2.4.

TABLE 2.4 BULB LIFETIMES IN HOURS, ORDERED DATA

Brand A		Brand B	
741	1026	772	1289
839	1033	941	1304
844	1087	1041	1309
869	1091	1073	1320
893	1125	1135	1322
897	1140	1141	1368
940	1149	1169	1381
973	1160	1204	1404
998	1167	1243	1462
1003	1242	1252	1471
1009	1302	1261	1575
1013	1417	1275	1653
1026		1278	

Now we can see more clearly that the Brand B bulbs tend to last longer. Furthermore, we can now pick out a number frequently used to characterize such sets of numbers.

If a set of values contains an odd number of values (as the two sets of our example do), its median is the middle value of the set, after the set has been arranged in order. In the two sets of our example, each of which contains 25 values, the medians are the thirteenth values—1026 hours for Brand A and 1278 hours for Brand B. If the median value occurs only once, as in the Brand B set, then an equal number of values of the set fall above and below it, 12 above and 12 below in this case. In the Brand A set the median value occurs twice, with the result that 12 values fall below the median, and another subset of 12 are greater than or equal to the median (only 11 being strictly greater).

For a set containing an even number of values, the median is defined as the average of the two middle values after the set has been rearranged in order. Thus the median of the set (2, 3, 7, 10, 11, 12) is $\frac{1}{2}(7 + 10) = 8.5$.

Summarizing our results, we have

	Median	Mean
Brand A	1026	1039
Brand B	1278	1266

We observe that the mean and median are fairly close in both cases, the mean falling above the median in the Brand A case and below in the Brand B case. In the Brand A case 15 values fall below the mean and 10 above; the mean does not have the property that the median does, of dividing the set into equal-numbered subsets of larger and smaller values.

The mean and the median are both measures of central tendency, numbers which may be useful in characterizing the typical value of the numbers in the set. Here it is hard to say which is a better indicator of central tendency, as their values are close compared to the spread of the data. Often, however, the mean and median differ considerably. Then it is a matter of judgment which is the better indicator of central tendency.

Questions

1. Suppose the members of the Central High School basketball team have heights as follows:

6'0"
6'3"
6'2"
5'10"
6'0"

6'6"
5'9"
6'1"
6'2"
5'11"

6'5"
5'6"
6'2"
6'9"
6'2"

(a) What is the median height?

(b) What is the average height? (See if you can devise a shortcut for this calculation.)

2. The last killing spring frost in a certain locality occurred on the following dates:

1961: April 21
1962: April 10
1963: May 10
1964: April 24
1965: April 17
1966: May 31
1967: May 1
1968: March 20
1969: April 20
1970: April 14

(a) What is the median of these dates?

(b) What is the average?

(c) Based on this data, what can you say about the recommended date for setting out tomato plants?

3. Is it easier to compute the median or the average

(a) in an unordered list of 10 numbers?

(b) in an unordered list of 1000 numbers?

(c) in an ordered list of 1000 numbers?

4. A sample of 20 members of the class of 1950 of Old Ivy University have annual salaries as given below:

$ 9,000
9,200
9,500
10,000
11,000
11,250
11,500
12,200
12,500
13,000
$ 13,500
14,500
15,000
16,500
18,500
20,000
26,500
39,500
85,000
120,000

(a) What is the median salary?

(b) What is the mean salary?

(c) Does the median or the mean better characterize the income of the members of the group?

(c) Should the median or the mean be used to plan fund-raising goals?

5. The median is preferred to the mean as a measure of central tendency when one suspects that irresponsible answers are present in the data.

Suppose 20 students are asked to estimate how much time they spend on homework. Their answers, listed in increasing magnitude for easy visualization, are:

Time Spent on Homework (in hours per week)	
-2	10
0	10
5	11
5	12
7	12.5
7.5	14
8	14
8	16
10	50
10	200

(a) What is the median? How much would it be likely to change if the obviously irresponsible answers were replaced by responsible ones?

(b) What is an approximate value for the mean? (Can you think of a quick way of estimating?)

(c) Why is the mean so much more sensitive to the irresponsible answers than the median?

2.6 Histograms and Frequency Distributions

A pictorial presentation of the bulb-lifetime data is possible if we classify it into intervals. Since the bulb lifetimes range from 741 to 1653 hours, let us take 10 class intervals, with boundaries as shown for Brand A in Table 2.5. It is then easy, starting with the raw data as given in Table 2.3, to count the occurrences in each interval by making hash marks as shown in Table 2.5. The resulting numbers of occurrences in each class are called frequencies. Thus we learn, for example, that five Brand A bulb lifetimes fell in the class interval of 800 - 899 hours; the frequency for that class is thus five.

TABLE 2.5. CLASSIFICATION OF BULB-LIFETIME DATA INTO INTERVALS (BRAND A)

Class Boundaries, Hours	Count of Occurrences	Frequency	Relative Frequencies	Class Marks
700 - 799	|	1	0.04	749.5
800 - 899	𝍸	5	0.20	849.5
900 - 999	|||	3	0.12	949.5
1000 - 1099	𝍸 |||	8	0.32	1049.5
1100 - 1199	𝍸	5	0.20	1149.5
1200 - 1299	|	1	0.04	1249.5
1300 - 1399	|	1	0.04	1349.5
1400 - 1499	|	1	0.04	1449.5
1500 - 1599		0	0.00	1549.5
1600 - 1699		0	0.00	1649.5

In Fig. 2.1 the frequencies have been pictured in a histogram. Frequencies are plotted vertically, and bulb lifetimes horizontally. Bars are drawn on the histogram, whose width is the class-interval width and whose height corresponds to the frequency of occurrences in the class. Thus, for the Brand A histogram, the bar for the 700 - 799 interval has height 1 corresponding to one occurrence in that class interval, etc.

The result is a display of the data that allows one to assess its nature more readily than by inspection of a column of numbers. Note the relationship between area on the histogram and number of occurrences; the ratio of the area of any bar to the total shaded area equals the relative frequency of that class (the fourth column in Table 2.5).

If one wishes to estimate the mean from the histogram, it is best to assign to all occurrences in a given class a value midway between the class

Fig. 2.1

boundaries. These values, called class marks, are given in the last column of Table 2.5. Denoting the frequencies by f_k and the class marks by c_k, the resulting approximate mean is given by

$$\overline{c} = \frac{1}{N}\sum_{k=1}^{n} f_k\, c_k \tag{2}$$

where $N = \sum_{k=1}^{n} f_k$ is the total number of occurrences and n the number of classes. For the Brand A mean lifetime this yields 1041.5 hours, very close to the true mean of 1039 hours.

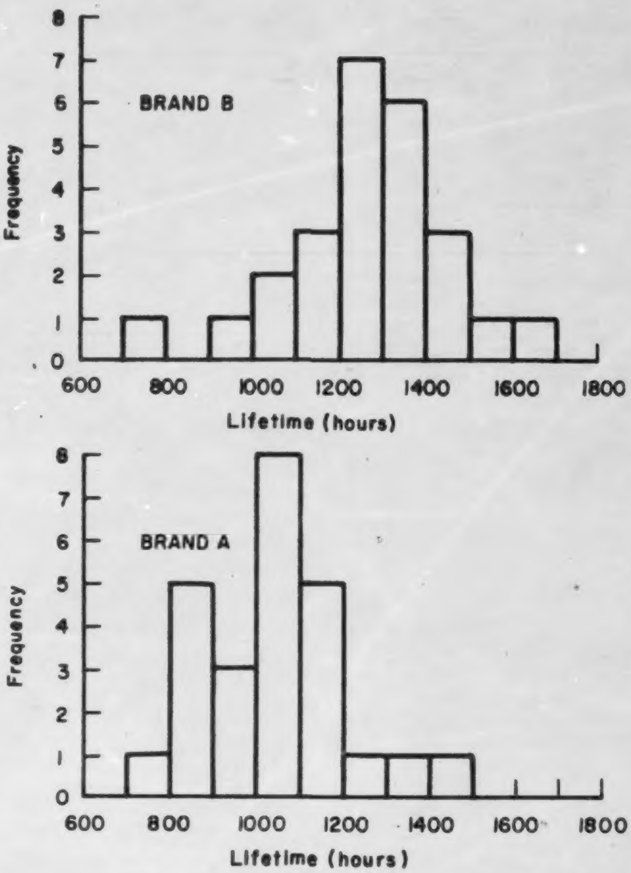


Fig. 2.1

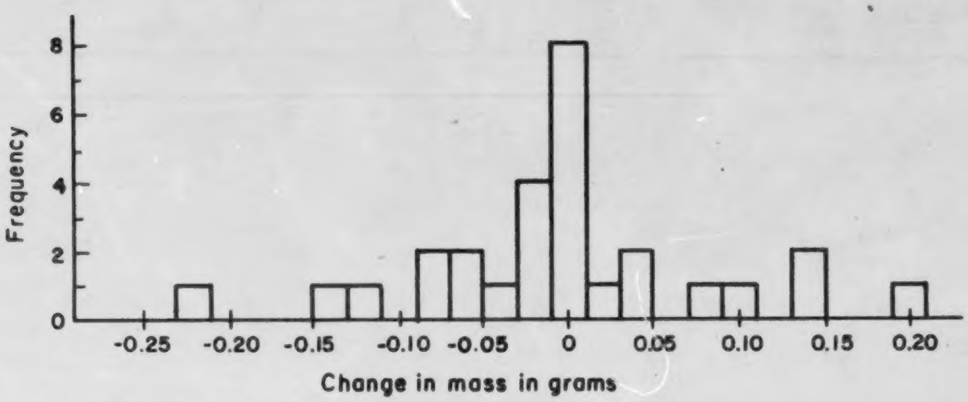
In careful work the class marks should be taken at the midpoints of the intervals. If in this example the class marks had been taken at 750 hours, 850 hours, etc. a constant bias in Equation (2) would have resulted, which, however, in this example would not be significant.

The discrepancy between the true mean and the approximate value given by Equation (2) is, of course, the result of the information that is lost in classifying the data into class intervals. If the intervals are excessively wide (so that there are only a few of them) this loss of information becomes serious. On the other hand, if the intervals are very narrow, there will be a large number of them and the computations become unnecessarily tedious. Usually a good compromise is a total of 10 to 20 class intervals. An exception is when a few of the values are far removed from a central cluster of values (as in the annual salary data in Question 4 of the preceding section); in such cases more than 20 intervals may be desirable.

Questions

1. Classify the Brand B Bulb lifetime data, from Table 2.3, into intervals by constructing a table of the form of Table 2.5. Verify that the Brand B histogram of Fig. 2.1 is correct.
2. On the basis of an intuitive visual inspection of the histograms of Fig. 2.1, mark the lifetime value that seems to you to characterize best the central tendency of the data. Now mark in the mean and the median values. Do these do well as indicators of central tendency for these examples?
3. In the light of the histograms of Fig. 2.1, can you make a final conclusion as to whether Brand A or Brand B is definitely better? (Note: the types of such conclusions that are possible, and the manner in which they may be drawn, is the concern of the field of statistics.)
4. Write the formula to estimate the Brand B lifetime mean from the frequency data you constructed in Question 1. If you have access to a desk computer, evaluate this approximation, and compare the result with the true value.
5. Suggest a quick way of estimating the median of data presented in histogram form. Estimate thereby the Brand A and Brand B bulb-lifetime data medians. What is the uncertainty associated with your method? What were the actual errors in your estimates?
6. Suppose you wished to present, in histogram form, data on the weights of individuals, in say, the entering freshman class of a certain college. Suppose the weights range from 96 to 234 pounds, and are reported to the nearest pound.
   (a) What class boundaries would it make sense to use, and how many classes would this yield?
   (b) What class marks correspond to the class boundaries you chose?
7. Explain why the median divides a histogram into two equal areas.

8. A piece of ice is massed before and after melting by 28 students. The resulting mass-change data are shown in the following histogram:



(a) Estimate the median of these data.

(b) Estimate the average.

(c) Why would evaporation tend to produce a negative bias in the mean, while massing errors would tend to produce fluctuations equally in the positive and negative directions? Which of these sources of error seems to be more important?

(d) What conclusions can be drawn from the aggregate of 28 trials of the experiment? Could such a conclusion be drawn from a single experiment?

(e) Is there reason to suspect from the data that some students have better laboratory technique than others? Explain.

2.7 Measures of the Spread of Data

Figure 2.2 shows, in histogram form, three sets of data all with the mean $\overline{x} = 9.9$. Although they have the same mean, these sets of data are clearly of different character; they are progressively more and more spread out. The mean, being a measure of central tendency, is of no help in describing the spread. How can we measure the extent of the spread of a set of data? Our approach will be to consider deviations from the mean and to apply an averaging process to these deviations.

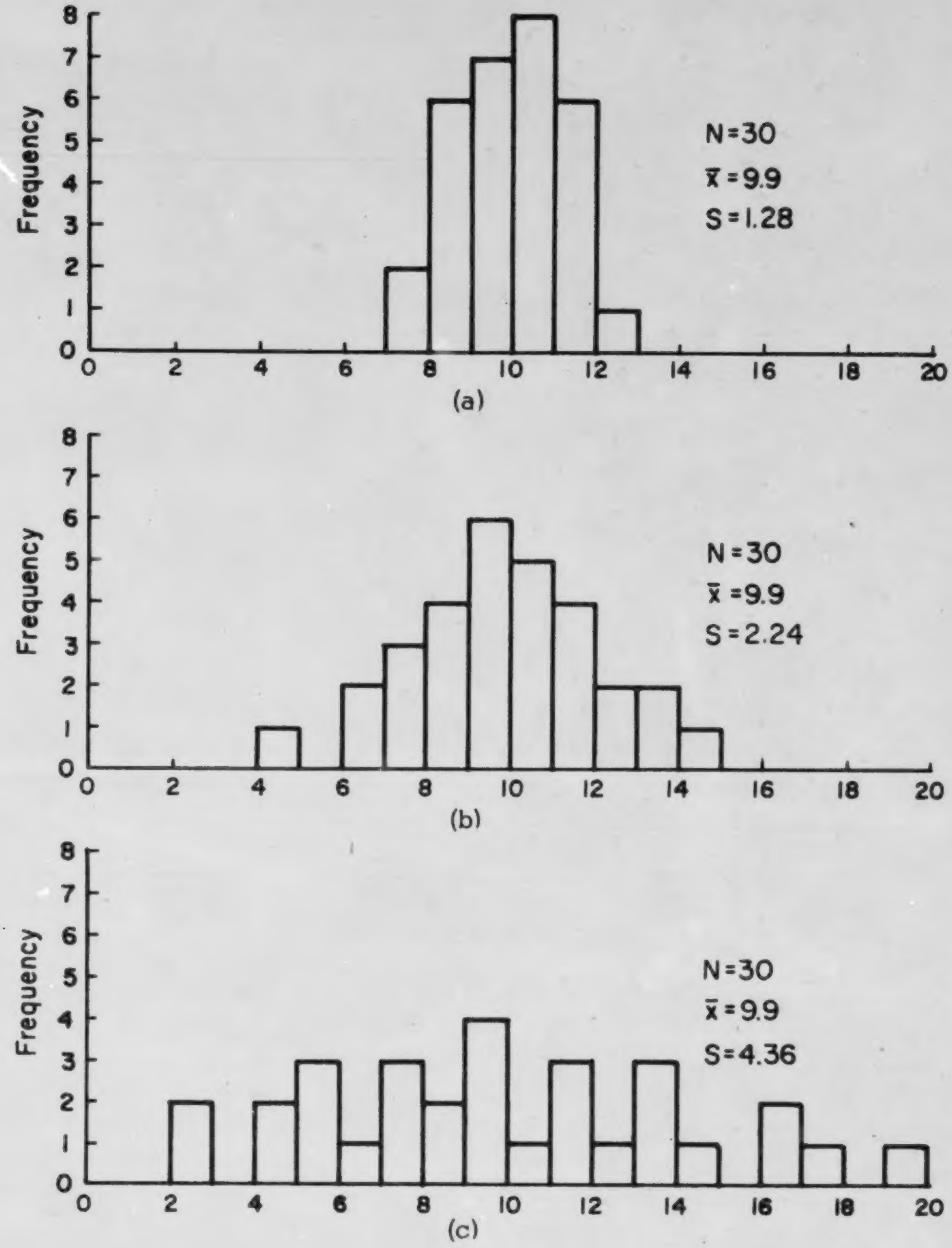

Frequency
0
1
2
3
4
5
6
7
8
0
2
4
6
8
10
12
14
16
18
20
N=30
x̄ =9.9
S=1.28
(a)
N=30
x̄ =9.9
S=2.24
(b)
N=30
x̄ =9.9
S=4.36
(c)

Fig. 2.2

Let us take an example with a comparatively small number of data, so that the computations will not be unduly long. (Fortunately, automatic computers are very well adapted to carrying out the types of calculations we shall describe and you will learn in Chapter 5 how to use a computer to handle larger and more realistic problems with comparative ease.

Suppose that, over the course of a year and under various driving conditions, you make 10 measurements of the gas mileage of your car, with the following results shown in Table 2.6 (these data were taken for a 1962 Volkswagen):

TABLE 2.6

GAS MILEAGE (mi/gal)

25.7	30.1
31.8	28.7
24.7	28.6
25.8	27.1
28.5	31.0

The processing of these data so as to measure the spread is shown in systematic form in Table 2.7

TABLE 2.7

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	25.7	-2.5	6.25
2	31.8	3.6	12.96
3	24.7	-3.5	12.25
4	25.8	-2.4	5.76
5	28.5	0.3	0.09
6	30.1	1.9	3.61
7	28.7	0.5	0.25
8	28.6	0.4	0.16
9	27.1	-1.1	1.21
10	31.0	2.8	7.84
Column Sums	282.0		50.38

The values are listed in the second column, and by summing this column and dividing by N = 10, we learn that the mean is $\overline{x} = 28.2$ mi/gal.

In the third column the deviations from the mean $x_i = \overline{x}$ are listed. Of course, some of these are positive and some negative, but their squares, listed in the fourth column, are all positive.

The mean of the squared deviations is called the variance. By summing the fourth column in Table 2.7 and dividing by N = 10, we obtain the value 5.04 for the variance. The variance is always non-negative, being computed by summing squares, so it is reasonable to denote it by s^2 as is the usual custom. The general formula for the variance is*

$$s^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2 \tag{3}$$

The square root s of the variance is called the root-mean-square deviation (rms deviation); or the standard deviation. From Equation (3) we see that s always has the same units as the original data. In our example s has the value $\sqrt{5.04} = 2.24$ miles per gallon.

In Table 2.7 we have carried "guard digits," even though they are often not significant, and have rounded off only at the end. In hand calculations, especially when no desk computer is available, to save time one often avoids carrying non-significant digits. However, automatic computers normally carry many places at no additional expense in labor. This is desirable because it prevents contamination of the final answer by round-off errors introduced during the calculations.

The dropping of non-significant digits, as a technique for keeping track of uncertainties, is too crude to be of much use in long calculations like those we have just done. The uncertainty is best estimated here by making small changes of, say, ±0.1 in the input data and recalculating to see the effect on the final results. If this is done for the above calculation, the final value of s is found to range from about 2.2 to 2.3, so it becomes apparent that the answer should be rounded to two digits.

---

*Some authors use (N-1) rather than N in the denominator, for a technical reason that need not concern us here.

It is apparent then that the standard deviation has the property that a measure of spread must have — namely that it is small when the data are concentrated about the mean, and large when the data are spread out. For when any value x_i is far removed from the mean $\overline{x}$, the corresponding term $(x_i - \overline{x})^2$ in Equation (3) makes a large contribution to the value of s^2. Consider the fourth column of Table 2.7, containing the values $(x_i - \overline{x})^2$ which are summed in computing s^2. We note that most of the contribution to s^2 corresponds to the value far removed from $\overline{x}$. In fact, the highest and lowest values, x_2 and x_3, alone account for over half of the value of s^2 in this example.

By modifying Equation (3) we can derive a short method for computing the standard deviation that is usually preferred to the method used in Table 2.7. Expanding Equation (3) we have

$$s^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i^2 - 2x_i\overline{x} + \overline{x}^2)$$

$$= \frac{1}{N} \sum_{i=1}^{N} x_i^2 + \frac{1}{N} \sum_{i=1}^{N} (-2x_i\overline{x}) + \frac{1}{N} \sum_{i=1}^{N} \overline{x}^2 \qquad (4)$$

The first term in Equation (4) is the average value of x_i^2, which we shall denote by $\overline{x^2}$:

$$\overline{x^2} = \frac{1}{N} \sum_{i=1}^{N} x_i^2$$

The second term in Equation (4) is a sum every term of which contains the constant factor $(-2\overline{x})$. Therefore $(-2\overline{x})$ may be factored out to yield

$$\frac{1}{N} \sum_{i=1}^{N} (-2x_i\overline{x}) = (-2\overline{x}) \frac{1}{N} \sum_{i=1}^{N} x_i$$

But we recognize this as $(-2\overline{x})$ times the mean $\overline{x}$, so this term equals $-2\overline{x}^2$.

The third and last term in Equation (4) is the sum as $\underline{i}$ ranges from 1 to N of the constant $\overline{x}^2$. Thus it equals

$$\frac{1}{N}\underbrace{(\overline{x}^2 + \overline{x}^2 + \ldots + \overline{x}^2)}_{\text{N times}} = \frac{1}{N}(N\,\overline{x}^2) = \overline{x}^2$$

Collecting these results together, we see that Equation (4) becomes

$$s^2 = \overline{x^2} - 2\overline{x}^2 + \overline{x}^2$$

$$s^2 = \overline{x^2} - \overline{x}^2 \tag{5}$$

This states that the variance is the mean of the squares minus the square of the mean.

Let us calculate $\underline{s}$ for our mileage data by this so-called "short" method using Equation (5). Table 2.8 lists the values of $x_i{}^2$ and we see that their sum is 8002.78. Dividing this by N = 10 yields $\overline{x^2} = 800.28$. If we subtract from this $\overline{x}^2 = (28.2)^2 = 795.24$, we get $s^2 = 5.04$, in agreement with our previous calculation.

TABLE 2.8

x_i	$x_i{}^2$
25.7	660.49
31.8	1011.24
24.7	610.09
25.8	665.64
28.5	812.25
30.1	906.01
28.7	823.69
28.6	817.96
27.1	734.41
31.0	961.00
282.0	8002.78

If the data are available only in histogram form, then in order to calculate approximate values of the mean and standard deviation, we assign to each value its class mark. The resulting calculation of s, for the data pictured in Fig. 2.2(b), is shown in Table 2.9. Note that the mean and the mean of the squares are weighted averages of the class marks c_i and their squares c_i^2, the weight factors f_i/N being the fraction of the total number of values in each class.

This type of calculation is also useful in the case of data which are "naturally classified" —that is, data which by their nature can take on only a relatively small number of discrete values. (For example, sample family sizes would be naturally classified data.)

TABLE 2.9

CALCULATION OF MEAN AND STANDARD DEVIATION OF CLASSIFIED DATA

i	Interval	Frequency f_i	Class Mark c_i	c_i^2	f_ic_i	$f_ic_i^2$
1	4 - 5	1	4.5	20.25	4.5	20.25
2	5 - 6	0	5.5	30.25	0.0	0.0
3	6 - 7	2	6.5	42.25	13.0	84.50
4	7 - 8	3	7.5	56.25	22.5	168.75
5	8 - 9	4	8.5	72.25	34.0	289.00
6	9 - 10	6	9.5	90.25	57.0	541.50
7	10 - 11	5	10.5	110.25	52.5	551.25
8	11 - 12	4	11.5	132.25	46.0	529.00
9	12 - 13	2	12.5	156.25	25.0	312.50
10	13 - 14	2	13.5	182.25	27.0	364.50
11	14 - 15	1	14.5	210.25	14.5	210.25
Column Sums		30			296.0	3071.50

$$\bar{x} = \frac{296.0}{30} = 9.87$$

$$s^2 = \frac{3071.50}{30} - (9.87)^2 = 5.02$$

Questions

1. Why is the sum of the numbers in the third column of Table 2.7 zero?
2. Estimate how the means and the standard deviations of the two distributions A and B in Fig. 2.1 compare.
3. Hypothetical meanings for the three sets of data in Fig. 2.2 are given below. In each case discuss briefly the significance of the varying degree of spread, and state which set best fits the given meaning.

   The data sets represent

   (a) test scores of sample groups of students on three alternative tests covering the same material;

   (b) failure loads of samples of three different types of sash cord to be used inside window frames;

   (c) sample lifetimes of three different types of automobile batteries;

   (d) trial shot-put distances of the members of three different track teams.
4. (a) In Table 2.7 approximate the standard deviation s by neglecting all values of $(x_i - \bar{x})^2$, except the four largest (that is, replacing the other six values of $(x_i - \bar{x})^2$ by zero). Compare with the exact value of s.

   (b) Do the same using only the two largest values of $(x_i - \bar{x})^2$. What point is this question trying to illustrate?
5. Ten students achieve the following scores on a test:

   8, 5, 7, 8, 6, 9, 4, 8, 2, 7

   (a) Draw a histogram for these data, and see if you can guess what the mean and standard deviations are.

   (b) Calculate the mean and standard deviations of these test scores by computing the mean squared deviation in the form given by Equation (3). How close were your guesses? Which values contribute most heavily to the standard deviation?

(c) Recalculate the standard deviation by the short method by computing the sum of the squares of the data and then using Equation (5). Compare this with the value you found in part (b). What are the relative merits of the two computational approaches?

6. Calculate the mean and standard deviation of the data of set (a) given in histogram form in Fig. 2.2.

7. Figure 2.3 shows two sets of data. Set (a) might represent data on the daily number of customers entering a certain store. Let us suppose that set (b) represents age at the time of marriage.

(a) These two data sets differ in a quality that is not directly related to their central tendency or spread. It is apparent from the shape of their distribution curves. Try to describe this quality, called "skewness," in words.

(b) In a set of data that is "skewed," as in Fig. 2.3(b), is the median displaced from the mean? If so, in which direction and why?

8. Consider sets of data of the various sorts listed below. In each case state whether you would expect the data set to be unskewed, as in Fig. 2.3(a), or skewed, as in Fig. 2.3(b), and why.

(a) Ages at which people contract mumps.

(b) Height of army recruits.

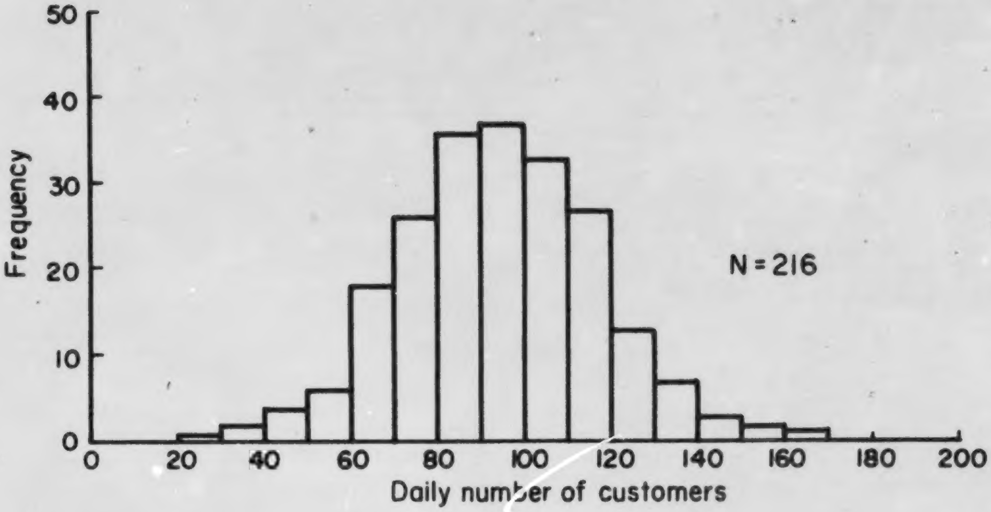
(c) Wealth of adults.

(d) Weight of new dimes.

(e) Weight of old dimes.

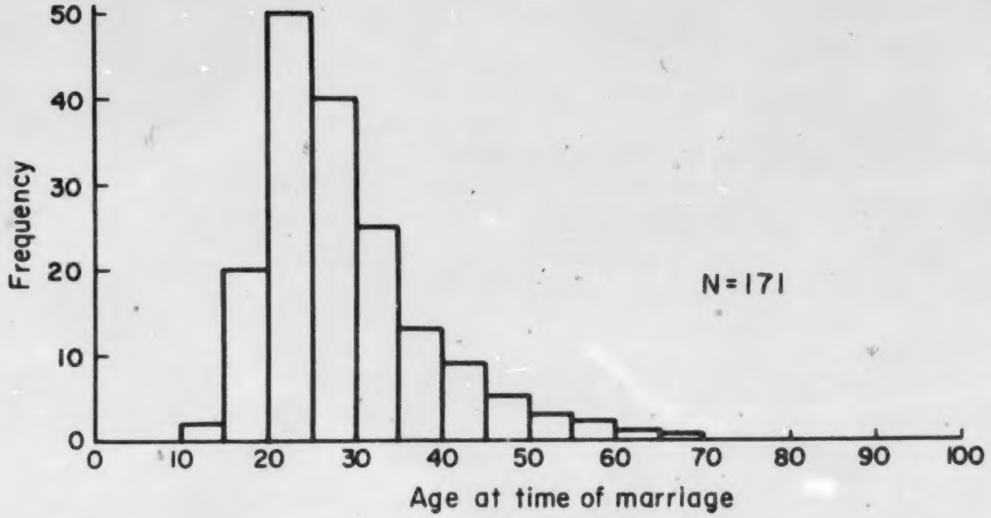
(f) Attendance at New York Mets baseball games.

(g) Number of home runs hit by members of the New York Mets in 1973.



50
40
30
20
10
0
Frequency
N = 216
0
20
40
60
80
100
120
140
160
180
200
Daily number of customers

(a)



50
40
30
20
10
0
Frequency
N = 171
0
10
20
30
40
50
60
70
80
90
100
Age at time of marriage

(b)

Fig. 2.3

9. In the 1960 U.S. census, records were made of the number of children of women in the age range 40-44 years, with the following results:

TABLE 2.10

No. of Children	Proportion of Total Women in Sample	No. of Children	Proportion of Total Women in Sample
0	0.141	7	0.019
1	0.172	8	0.012
2	0.262	9	0.007
3	0.182	10	0.005
4	0.105	11	0.003
5	0.056	More than 11	0.005
6	0.031		

(a) Comment on the degree of skewness of this data set.

(b) Calculate the mean number of children of such women.

(c) Estimate the uncertainty in your answer to part (b) due to the 0.005 in the "more than 11" category. What assumption did you make about this category in answering part (b)?

(d) Determine the median number of children. When might we use this as a measure of central tendency? When would we prefer the mean?

(e) If one were interested in population growth, why might the above data be preferable to, say, data on U. S. family sizes? Why do you suppose women in the age group 40-44 years were chosen rather than a younger or older age group?

10. In order to describe and compare data sets, it is sometimes useful to employ the idea of percentile, a generalization of the idea of median. The median is also called the 50 percentile, meaning that 50 per cent of the values fall below it. Correspondingly, the 25 percentile is a value below which 25 per cent of the values fall, the 90 percentile is a value below which 90 per cent of the values fall, etc.

(a) How might the idea of percentile be used to obtain a measure of spread?

(b) Use the method you propose to compare the extent of spread of the two sets of bulb lifetime data given in Table 2.4.

(c) What are the pros and cons of this measure of spread versus calculating the standard deviation?

11. In 1968 the American League winning baseball scores were as shown in Table 2.11 (source: Official Baseball Guide for 1969, published by Sporting News, St. Louis).

TABLE 2.11

Score	No. of Games		Score	No. of Games
1	38		9	21
2	101		10	21
3	131		11	12
4	159		12	10
5	110		13	6
6	82		14	1
7	73		15	0
8	44		16	1
			Total No. Games:	810

(a) Comment on the skewness of this data set.

(b) Give the 10, 25, 50, 75, 90, and 95 percentile scores.

(c) Calculate the mean and standard deviations of the 1968 winning scores. (Do this part only if you have access to a desk computer or equivalent.) Comment on the displacement between the mean and the median.

12. Most examples of sets of data we have discussed may be characterized as having a single "hump," containing the majority of the data, with tails on either side. Sometimes data does not have such "nice" regular behavior. For example, the New England Board of Higher Education, in Facts about New England Colleges, Universities and Institutes, 1971-72, reported tuition of such institutions in Maine (for state residents) as follows:

TABLE 2.12

$ 865	$2350	$ 400
3525	1020	400
1100	445	400
1210	1850	400
2795	2000	400
2660	1675	550
247	287	450
1650	1600	550
700	1700	400
		1530

(a) Plot these data in a histogram, using an interval of $200.

(b) Are these sets of data well characterized by giving the mean and standard deviation, or would more have to be specified to convey their general characteristics?

(c) Describe in words the nature of these data. See if you can think of any possible reasons for any of their characteristics.

13. According to the 1969 World Almanac and Book of Facts, published by the Boston Herald Traveler, the 1968 winning college football scores for 1173 games were distributed as follows. (These include the scores of the winners of all games and the tie scores in tied games:

TABLE 2.13

Score	Frequency	Score	Frequency	Score	Frequency	Score	Frequency
0	5	20	48	40	17	60	3
1	0	21	70	41	22	61	1
2	0	22	18	42	34	62	2
3	2	23	32	43	9	63	4
4	0	24	45	44	10	64	1
5	0	25	14	45	13	65	2
6	7	26	29	46	14	66	1
7	25	27	51	47	20	67	0
8	3	28	76	48	16	68	6
9	7	29	18	49	12	69	2
10	28	30	31	50	9	70	0
11	1	31	48	51	2	71	1
12	9	32	17	52	4	72	0
13	26	33	20	53	3	73	0
14	46	34	36	54	2	74	0
15	3	35	54	55	7	75	0
16	26	36	10	56	5	76	1
17	48	37	23	57	3	77	1
18	12	38	21	58	9	•	•
19	13	39	10	59	4	•	•
						•	•
						100	1

This is an example of a data set exhibiting a good deal of "fine structure"—local peaks and valleys, etc. Describe some of these features, and explain why these data are so much more complicated than the winning baseball scores of Question 11. Is there any similarity with the baseball scores data?

14. Sometimes data is tabulated in unequal interval sizes. Where age is concerned (Table 2.14) unequal interval sizes are common practice.

TABLE 2.14

Age	Native Born Population	Death Rate per 1000
Less than 1	3,414,000	23.4
1 - 4	13,380,000	0.9
5 - 14	29,505,000	0.4
15 - 24	20,091,000	1.0
25 - 34	18,842,000	1.2
35 - 44	20,004,000	2.6
45 - 64	28,561,000	10.6
65 - 74	7,699,000	36.1
75 - 84	3,181,000	87.2
85 and over	625,000	210.6

As can be seen from the death rates, the risk of dying in the first year of life is very different from immediately subsequent years, thus it makes sense to consider that age range separately. Construct a histogram of the native born population data using the age intervals given. (Hint: which should be proportional to the frequency, the height of the bars or their area?) Comment on the skewness of these sets of data.

Chapter 3. NUMERICAL CALCULATIONS

3.1 Large and Small Number Calculations

In Chapter 1 we introduced the use of powers-of-ten notation. This way of expressing a number is also called exponential notation because of the use of exponents of 10. Exponential notation is very useful in performing calculations with both large and small numbers.

To make calculations involving large numbers expressed in exponential notation, recall that $10^a \times 10^b = 10^{a+b}$. Thus, for example

$$(15 \times 10^6) \times (3.0 \times 10^5) = (15 \times 3.0) \times (10^6 \times 10^5) = 45 \times 10^{11}$$

In additions and subtractions the numbers given in exponential notation must be re-expressed, if necessary, so that the exponents are the same. For example,

$$\begin{aligned} 5.32 \times 10^3 + 2.11 \times 10^2 &= 5.32 \times 10^3 + 0.211 \times 10^3 \\ &= (5.32 + 0.211) \times 10^3 \\ &= 5.53 \times 10^3 \end{aligned}$$

The more complex a large-number calculation is, the more useful exponential notation becomes. How much water is used by New York City in a year? It has been estimated that a typical city uses about 1.4×10^2 gallons of a water a day for each of its residents. According to the 1970 census, the population of New York was 7.89×10^6. Therefore, the city used, in 1970, about $(8.0 \times 10^6)(1.4 \times 10^2)$ gallons each day, or $(8.0 \times 10^6)(1.4 \times 10^2)(3.7 \times 10^2)$ gallons every year. This is approximately 4×10^{11} gallons per year.

Calculations involving very small numbers are also often best done using exponential notation. For example, given that 1.00×10^3 g of copper contains 9.4×10^{24} atoms, what is the mass of one atom? It is given by the mass of the sample divided by the number of atoms in the sample.

$$\text{mass of one atom} = \frac{1.00 \times 10^3 \text{ g}}{9.4 \times 10^{24}} = 0.106 \times \frac{10^3}{10^{24}} \text{ g}$$

The rule for dividing powers-of-ten is $\frac{10^a}{10^b} = 10^{a-b}$ (easily verified by multiplying both sides by 10^b). Thus we have

$$\text{mass of one atom} = 0.106 \times 10^{3-24}\ \text{g}$$
$$= 0.106 \times 10^{-21} \text{ or } 1.06 \times 10^{-22}\ \text{g}$$

Questions

1. What is 10^0? Justify your answer.

2. Which of the following is correct? For the ones that are incorrect, explain how the person giving the answer went wrong. Change the right-hand side of the equation to correct the error.

   (a) $10^6 \times 10^0 = 10^0$

   (b) $10^{-3} \times 10^2 = 10^{-1}$

   (c) $10^{-3} \times 10^2 = 10^{-6}$

   (d) 10^{-6} is larger than 10^{-3}

   (e) $10^{-4} \times 10^{-3} = 10^{12}$

3. In each of the following lists, indicate the numbers that are equal to each other.

   (a) 0.003
   3×10^{-2}
   0.3×10^{-2}
   $3 \times \frac{1}{10^2}$

   (b) 0.000028
   28×10^{-5}
   2.8×10^{-5}
   0.28×10^{-5}

4. Using exponential notation, calculate answers to the following:

   (a) $2300 \times 4600 \times 120$

   (b) $\frac{4700 \times 0.32 \times 5000}{13 \times 0.0046}$

5. In the text we found an approximate value for the annual water consumption of New York City. How many square kilometers of watershed are needed to supply the New York City reservoirs? The annual rainfall in the New York area is about 1.0 m. To visualize the amount of water falling on a square kilometer in one year, think of a rectangular volume whose base is a square with sides measuring 1.00 kilometer and whose height is 1.0 m. Assume that half of the rain that falls on the watershed gets to the reservoirs.

6. The total land area of the United States is about 2×10^7 square kilometers. If the land were distributed evenly among the population (about 2×10^8), approximately how much land would each person receive?

7. The decimal expansion of π is 3.141592653589793.... . It has been calculated by computer to 100,000 decimal places. This calculation took 8 hours and 43 minutes of computer time working at an average speed of over 100,000 arithmetic operations (multiplications or additions) a second. It has been estimated that the same job using a desk calculator would take about 30,000 years.

   (a) Approximately how many arithmetic operations did the computer do altogether?

   (b) How many times longer would it take to compute π to 100,000 decimal places by desk calculator than by computer?

8. About how many times does an automobile tire (outside diameter about 75 centimeters) rotate in traveling 10,000 kilometers? If a centimeter of tread is worn off in going this distance, about how much thickness of tread is worn off during one rotation, on the average?

9. The speed of light is 3.00×10^8 meters per second. How long does it take light to travel 10.0 meters?

10. An ordinary land snail can move with a speed of 8×10^{-3} kilometers (5×10^{-3} miles) per hour.

    (a) At this rate, crawling steadily, how long would it take such a snail to cross the United States?

    (b) If the average life span of a snail is five years, how many generations would this journey represent?

3.2 Estimation

Very often, one is interested in getting the approximate magnitude of some quantity when the values of quantities to be used in the calculation are not available. Sometimes a rough estimate of the unknown values can be obtained using related known information.

For example, suppose you are interested in estimating the total number of miles traveled by private cars in the United States each year. If you know approximately how many cars there are in the United States and how many miles each is driven during a year on the average, then you could multiply these two numbers together to get the answer. But you do not even have a rough idea of the number of cars. However, it might be reasonable to suppose that, very roughly, the average family has four people in it and owns one car. There are about 2×10^8 people in the United States and thus about $\frac{2 \times 10^8}{4} = 5 \times 10^7$ such four-person families. Hence there are about 5×10^7 cars. A typical yearly distance for a car to travel, from personal experience, might be about 10^4 miles. Thus the total distance traveled by cars in the United States each year is about $(5 \times 10^7) \times (10^4) = 5 \times 10^{11}$ miles.

How can we estimate the uncertainty in this answer? We might judge that one car per every two people is definitely higher than the true figure. Similarly we might judge that 25,000 miles is definitely high for the average yearly distance per automobile. This would imply that the total mileage is less that $10^8 \times 2.5 \times 10^4 = 2.5 \times 10^{12}$ miles. By similarly making low estimates we can deduce that the total mileage is probably greater than $(2 \times 10^7) \times (5 \times 10^3) = 10^{11}$ miles. In other words, the true figure is probably not more than 500 per cent more, nor 80 per cent less than our estimate of 5×10^{11} miles. We are not off by a factor larger than 5 or smaller than $\frac{1}{5}$.

Perhaps surprisingly, though, such crude answers are frequently adequate. That is to say, frequently we want to know only the order of magnitude of a very large quantity such as this. We can say here with assurance that the order of magnitude of the United States private car annual mileage is 5×10^{11} miles.

Sometimes a simple experiment helps one to arrive at a good estimate of some quantity. For example, about how many words are there in a book? To find out, you need to know the number of pages in the book and the average number of words on a page. To estimate the latter number, one might count the words in a line chosen at random, and then multiply it by the number of lines on one page.

Estimation and approximation are not synonymous. In an approximation the numbers are given and only the calculation is approximate. In an estimate one or more numbers entering into the calculation are approximated by an educated guess or very rough measurement.

Questions

1. About how many revolutions does the wheel of an automobile make in a trip from New York to Los Angeles?
2. Estimate the uncertainty in your answer to Question 1.
3. In estimating the number of words in a book, why might it be better to count the words in 10 lines and divide by 10 rather than counting the words in a single line as suggested in the text?
4. Estimate each of the following, and indicate how you arrived at your answer:
   (a) The total amount of gasoline consumed by automobiles in the United States each year.
   (b) The number of words in an encyclopedia.
   (c) The number of words in an average half-hour news broadcast.
   (d) The number of tin cans used in United States homes each year.
   (e) The volume of concrete in one mile of an interstate highway.
5. Estimate the uncertainty in each of your answers to Question 4. Express each uncertainty in both absolute and relative form. In which case is the result known only to within an order of magnitude?

6. Estimate the orders of magnitude of:
    - (a) The number of shingles on a shingled roof.
    - (b) The number of bricks in a brick house.
    - (c) The number of dwellings (including apartments) in your city.
    - (d) The number of classroom chairs in a given school or college.
7. Estimate the volume of a warehouse that would be needed to store a year's production of ping-pong balls in the United States. By how many orders of magnitude might your answer be off?
8. How many seconds are there in an average human lifetime?

3.3 First Order Approximations

Consider the following numbers: 1.039^2, 1.0056^3, or $\frac{1}{0.973}$. They have one property in common; they are the result of some operation with numbers which differ only slightly from 1. These numbers are just examples of a general class of numbers which can be written as $(1+\epsilon)^2$, $(1+\epsilon)^3$, and $\frac{1}{1+\epsilon}$ where the Greek letter ϵ is customarily used to indicate numbers whose absolute value is small compared to 1. In mathematical notation this condition is written as $|\epsilon| << 1$. The values of ϵ in the three examples are 3.9×10^{-2}, 5.6×10^{-3}, and -2.7×10^{-2} respectively.

In this section we wish to show that there exist useful ways of finding approximate values for expressions of the type $(1+\epsilon)^2$, $(1+\epsilon)^3$, and $\frac{1}{1+\epsilon}$ where $|\epsilon| << 1$.

Let us start with the first two expressions: In general

$$(1+\epsilon)^2 = 1 + 2\epsilon + \epsilon^2$$

and

$$(1+\epsilon)^3 = 1 + 3\epsilon + 3\epsilon^2 + \epsilon^3$$

For $|\epsilon| << 1$, the term proportional to ϵ^2 is much smaller than the term proportional to ϵ in both cases. For example, if $\epsilon \approx 10^{-2}$, then $\epsilon^2 \approx 10^{-4}$, and the term $\epsilon^3 \approx 10^{-6}$ is, of course, still smaller. Hence for $|\epsilon| << 1$

$$(1+\epsilon)^2 \approx 1 + 2\epsilon \tag{1}$$

and

$$(1+\epsilon)^3 \approx 1 + 3\epsilon \tag{2}$$

To see what error is involved in making these approximations we subtract the approximate expression from the exact one. In the first case the error is

$$(1 + \epsilon)^2 - (1 + 2\epsilon) = \epsilon^2$$

and in the second case

$$(1 + \epsilon)^3 - (1 + 3\epsilon) = 3\epsilon^2 + \epsilon^3 = \epsilon^2(3 + \epsilon)$$

The absolute value of the factor $(3 + \epsilon)$ has an upper bound for $|\epsilon| << 1$; we can state with certainty that under this condition $|3 + \epsilon| < 4$. Thus the error is never larger than $4\epsilon^2$. When the error in an approximation can be shown to be less than a constant times ϵ^2, we say that the approximation is the first order approximation in ϵ. Thus $1 + 2\epsilon$ is the first order approximation for $(1 + \epsilon)^2$ and $1 + 3\epsilon$ is the first order approximation for $(1 + \epsilon)^3$.

Now let us find the first order approximation to $\frac{1}{1 + \epsilon}$. By long division (or by adding "the well-chosen zero"*, $\epsilon - \epsilon$ in the numerator, twice) we find

$$\frac{1}{1 + \epsilon} = 1 - \epsilon + \frac{1}{1 + \epsilon}\epsilon^2 \qquad (3)$$

For $|\epsilon| << 1$, say $|\epsilon| \leq 0.1$ the absolute value $\left|\frac{1}{1 + \epsilon}\right| \leq \frac{1}{0.9} = \frac{10}{9}$. Hence if we approximate $\frac{1}{1 + \epsilon}$ by $1 - \epsilon$ the magnitude of the error is less than $\frac{10}{9}\epsilon^2$. Hence $1 - \epsilon$ is the first order approximation for $\frac{1}{1 + \epsilon}$.

How good any of these first order approximations are depends on the degree of accuracy required in the particular application. As long as the factor multiplying ϵ^2 is less than some known constant we can always estimate the error made in the first order approximation.

Questions

1. To appreciate the usefulness of the first order approximations evaluate the following expressions (i) to first order in ϵ and (ii) exactly:

   (a) $(1 + \epsilon)^2$ for $\epsilon = -0.007$

   (b) $(1 + \epsilon)^3$ for $\epsilon = 0.05$

   (c) $\frac{1}{1 + \epsilon}$ for $\epsilon = 0.011$

---

*See Appendix 2

2. Suppose the numbers given in Question 1 are physical numbers with an uncertainty of one unit in the last digit. Would you need to go beyond the first order approximation? Would the first order approximation suffice for $(1 + \epsilon)^3$ where $\epsilon = 0.4$? What is the relative error in this case?
3. Find the first order approximation for $(1 + \epsilon)^4$ and prove that it satisfies the condition that the error is less than ϵ^2 times a constant.
4. For $|\epsilon| << 1$ the number $1 + 5\epsilon$ is certainly an approximation for $(1 + \epsilon)^4$. (In fact it is a better approximation than $1 + 2\epsilon$.) Why does it not qualify as the first order approximation for $(1 + \epsilon)^4$?
5. Find the first order approximation in ϵ for $\frac{1 + 2\epsilon}{1 - 3\epsilon}$ for $|\epsilon| << 1$ and use it to calculate $\frac{1.04}{0.94}$.
6. Find the first order approximation for $\frac{1}{(1 + \epsilon)^2}$. Use it to calculate $\frac{1}{1.12^2}$

3.4 An Extension of First Order Approximations

In the preceding section the expressions we approximated involved numbers close to 1. Can we use similar approximations for expressions involving numbers close to a given number other than 1? For example, does our knowledge that $5^3 = 125$ help us to find 5.07^3? To put the question in a more general form does the knowledge of A^3 help us in finding an approximate value of $(A + a)^3$ where $|a| << |A|$?

Since $A + a = A(1 + \frac{a}{A})$, then $(A + a)^3 = A^3(1 + \frac{a}{A})^3$. From $|a| << |A|$ it follows that $\left|\frac{a}{A}\right| << 1$, thus the ratio $\frac{a}{A}$ now takes the place of ϵ in the preceding section, and we see that, to first order, $(A + a)^3 \approx A^3(1 + 3\frac{a}{A})$. We can use this approximation whenever $\left|\frac{a}{A}\right| << 1$, i.e., the relative difference between the two numbers A + a and a is much less than 1.

In applications the numbers A and A + a may be dimensional numbers and hence their values will depend on the units used (e.g., 2 meters or 200 cm). However the ratio $\frac{a}{A}$ is always a pure, dimensionless number and hence the condition $\left|\frac{a}{A}\right| << 1$ is independent of the units of A.

Questions

1. Using the first order approximations developed in the preceding section and the relations $(A + a)^2 = A^2(1 + \frac{a}{A})^2$ and $\frac{1}{A + a} = \frac{1}{A} \cdot \frac{1}{1 + \frac{a}{A}}$ calculate 5.15^2, 7.92^2, 100.3^2, $\frac{1}{10.2}$, and $\frac{1}{98.3}$.

2. What is the relative difference between the areas of two squares with sides 6.00 m and 6.24 m?

3. Suppose you wish to apply a first order approximation to calculate 10.2^3 using the value of 10^3, and 20.3^3 using the value of 20^3. In which case will the approximation be better? (Be sure to specify which criterion you are using for the quality of the approximation.)

3.5 Relative Uncertainties in Multiplication and Division

The first order approximations developed in the preceding sections for mathematical numbers can be applied directly to physical numbers that have small relative uncertainties. Consider a physical number A with an uncertainty $\pm a$. The square of this number will most likely lie between $(A + a)^2$ and $(A - a)^2$. If $|a| << A$, it will suffice to calculate the square to first order in $\frac{a}{A}$:

$$(A \pm a)^2 = [A(1 \pm \frac{a}{A})]^2 \approx A^2 \pm 2Aa = A^2(1 \pm 2\frac{a}{A})$$

Thus, for physical numbers with small relative uncertainties, the relative uncertainty in the square of the number is twice the relative uncertainty in the number itself.

Let us now extend this result to the product of two different positive physical numbers: $(A \pm a)(B \pm b)$. The product is most likely to be between $(A + a)(B + b)$ and $(A - a)(B - b)$.

Suppose that $\frac{a}{A} > \frac{b}{B}$. Then it is convenient to choose a number ϵ such that $\frac{a}{A} = k_1\epsilon$ and $\frac{b}{B} = k_2\epsilon$, where k_1 is of order 1 and k_2 is of order 1 or less. (For example, if $\frac{a}{A} = 0.027$ and $\frac{b}{B} = 0.005$, we may choose $\epsilon = 10^{-2}$, which makes $k_1 = 2.7$ and $k_2 = 0.5$.) Then

$$\begin{aligned}(A + a)(B + b) &= AB\left(1 + \frac{a}{A}\right)\left(1 + \frac{b}{B}\right) \\ &= AB(1 + k_1\epsilon)(1 + k_2\epsilon) \\ &= AB[1 + (k_1 + k_2)\epsilon + k_1k_2\epsilon^2]\end{aligned}$$

Hence, to first order in ϵ

$$\begin{aligned}(A + a)(B + b) &\approx AB[1 + (k_1 + k_2)\epsilon] \\ &= AB\left[1 + \left(\frac{a}{A} + \frac{b}{B}\right)\right]\end{aligned}$$

Following the same steps for the lower end of the interval yields

$$(A - a)(B - b) \approx AB\left[1 - \left(\frac{a}{A} + \frac{b}{B}\right)\right]$$

Hence

$$(A \pm a)(B \pm b) \approx AB\left[1 \pm \left(\frac{a}{A} + \frac{b}{B}\right)\right] \tag{4}$$

We see that in multiplication small relative errors add.

Using the first order approximation for reciprocals will show that small relative errors also add in the case of the division of two positive physical numbers. The ratio of $\frac{A \pm a}{B \pm b}$ is between $\frac{A + a}{B - b}$ and $\frac{A - a}{B + b}$. Again, setting $\frac{a}{A} = k_1\epsilon$ and $\frac{b}{B} = k_2\epsilon$, we have

$$\frac{A + a}{B - b} = \frac{A(1 + k_1\epsilon)}{B(1 - k_2\epsilon)} = \frac{A}{B}(1 + k_1\epsilon)\left(1 + k_2\epsilon + \frac{(k_2\epsilon)^2}{1 - k_2\epsilon}\right)$$

Multiplying out the right-hand side we get

$$\frac{A + a}{B - b} = \frac{A}{B}\left[1 + (k_1 + k_2)\epsilon + \left(k_1k_2 + \frac{(k_2)^2}{1 - k_2\epsilon} + \frac{k_1(k_2)^2\epsilon}{1 - k_2\epsilon}\right)\epsilon^2\right]$$

The coefficient of ϵ^2 in the last equation has an upper bound for $|\epsilon| << 1$. Hence to first order in ϵ

$$\frac{A + a}{B - b} \approx \frac{A}{B}[1 + (k_1 + k_2)\epsilon] = \frac{A}{B}\left(1 + \frac{a}{A} + \frac{b}{B}\right) \tag{5a}$$

Similarly

$$\begin{aligned}\frac{A - a}{B + b} &= \frac{A}{B}\frac{(1 - k_1\epsilon)}{(1 + k_2\epsilon)} \approx \frac{A}{B}(1 - k_1\epsilon)(1 - k_2\epsilon) \\ &\approx \frac{A}{B}\left[1 - \left(\frac{a}{A} + \frac{b}{B}\right)\right] \qquad (5b)\end{aligned}$$

This proves our claim.

Questions

1. For motion at constant speed, the distance traveled equals the product of speed and time. If the speed is measured within 3 per cent and the time is measured within 2 per cent, what is the relative uncertainty in the calculated distance?
2. Extend the proof given in the text for the relative uncertainty of a product of two physical numbers to a product of three numbers. Compare the special case where the three numbers are equal with the first order approximation for $(A + a)^3$.
3. The sides of a rectangular box are found to be 5.00 ±0.01 cm, 6.00 ±0.01 cm, and 2.00 ±0.01 cm.
   (a) What is the relative error in each dimension?
   (b) What is the relative error in the volume of the box?
   (c) Would your answer to part (b) be different if the dimensions of the box were 5.02 ±0.01 cm, 6.13 ±0.01 cm, and 1.92 ±0.01 cm?
4. The density of a substance is calculated by dividing the mass of a sample by its volume. Suppose the sample is a cube. The length of its side is measured to ±2 per cent; its mass is known to ±1 per cent.
   (a) What is the relative uncertainty in density?
   (b) Suppose you have to know the density to a higher accuracy. Would it be better to improve the measurement of the length of the side to give a relative error of ±1 per cent or reduce the relative error in the mass to 0.2 per cent?

3.6 Finding Square Roots by an Iterative Process

Square roots come up frequently in numerical calculations. Most square roots of mathematical numbers such as $\sqrt{2}$ cannot be written as exact decimal numbers. However, in this section we shall show how to calculate an approximation to any square root with as great an accuracy as desired.

A natural way to get an approximate value for $\sqrt{2}$ is to try possible values. Since $(1)^2 = 1$ and $(2)^2 = 4$, $\sqrt{2}$ must be between 1 and 2. We might therefore try 1.3 as a first crude approximation. To see how good our guess is, we divide 2 by this guess and get $\frac{2}{1.3} = 1.54$. Thus we know that $1.3 \times 1.54 = 2$. It is apparent that $1.3^2 < 2$, whereas $1.54^2 > 2$. Therefore $\sqrt{2}$ lies between 1.3 and 1.54. We now take a value halfway between these two values as our next approximation: $\frac{1.3 + 1.54}{2.0} = 1.42$. The difference between this approximation to $\sqrt{2}$ and the upper and lower bounds, 1.54 and 1.3, is 0.12, which means that the relative error is no greater than $\frac{0.12}{1.42} \approx 0.08$ or ±8 per cent.

We can repeat this procedure a second time and thereby reduce the difference between the upper and lower bound for $\sqrt{2}$. Dividing 2 by the last guess gives $\frac{2}{1.42} = 1.41$. Thus $1.41 < \sqrt{2} < 1.42$, and the average of these two values, 1.415, is not more than ± 0.5 per cent in error. This process can be repeated until the desired accuracy has been obtained. A process such as this, in which successively better approximations are obtained by repeating the same step, is called an iterative process.

An iterative process requires a way of starting the procedure and the existence of a clear instruction of how to proceed from step $\underline{k}$ to step $k + 1$. In the case of finding the square root of a positive number $\underline{N}$, we have, starting with the initial guess x_0

$$
\begin{aligned}
x_1 &= \frac{1}{2}(x_0 + \frac{N}{x_0}) \\
x_2 &= \frac{1}{2}(x_1 + \frac{N}{x_1}) \qquad (6) \\
x_3 &= \frac{1}{2}(x_2 + \frac{N}{x_2}) \\
&\vdots \\
x_{k+1} &= \frac{1}{2}(x_k + \frac{N}{x_k}), \qquad k = 0, 1, 2, \ldots
\end{aligned}
$$

This is called the iteration formula for this iterative process. In Chapter 5 we shall use it as an example of automatic computation by a computer.

Note that there is only a positive square root of a positive number, and by convention we mean this square root when we write $\sqrt{x}$. For example, $\sqrt{4} = 2$, not -2. To indicate the negative square root we must write $-\sqrt{x}$, and to indicate both positive and negative square roots we write $\pm\sqrt{x}$.

Questions

1. (a) Find $\sqrt{10}$ to within 0.5 per cent.

   (b) When p and q are non-negative numbers, $\sqrt{pq} = \sqrt{p} \times \sqrt{q}$. Use this relation to calculate $\sqrt{20}$, $\sqrt{200}$, and $\sqrt{2 \times 10^7}$.

2. Find $\sqrt{1/2}$ to within 1 per cent.

3. Find $\sqrt[4]{10}$ to within 1 per cent.

4. An iterative process is said to be self-correcting if it approaches some number even if a mistake is made at some point in the calculations. Is the square-root iterative process self-correcting? Explain.

5. Devise a method for finding cube roots, similar to the iterate method for finding square roots. (Hint: If x_0 were the cube root of N we would have $\frac{N}{{x_0}^2} = x_0$. But of course $\frac{N}{{x_0}^2}$ equals some other number y. Where must the cube root of N be with respect to x_0 and y? How would you calculate the next approximation x_1 ?)

3.7 The First Order Approximation for $\sqrt{1+\epsilon}$

The iterative process for finding square roots which we developed in the preceding section can be used to find a first order approximation for $N = \sqrt{1+\epsilon}$ where $|\epsilon| << 1$. We know that $\sqrt{1+\epsilon}$ must be close to 1, so we choose as our first guess $x_0 = 1$. The next step gives

$$x_1 = \frac{1}{2}\left(x_0 + \frac{N}{x_0}\right) = \frac{1}{2}\left(1 + \frac{1+\epsilon}{1}\right) = 1 + \frac{\epsilon}{2}$$

This tells us that

$$1 < \sqrt{1+\epsilon} < 1 + \frac{\epsilon}{2}$$

To be sure that $1 + \frac{\epsilon}{2}$ is the first order approximation for $\sqrt{1+\epsilon}$, we must show that $\sqrt{1+\epsilon} - (1 + \frac{\epsilon}{2}) <$ constant times ϵ^2. To do that we proceed to the next step in the iteration

$$x_2 = \frac{1}{2}\left(1 + \frac{\epsilon}{2} + \frac{1+\epsilon}{1+\epsilon/2}\right) \tag{7}$$

Applying long division to the third term gives

$$\frac{1 + \epsilon}{1 + \epsilon/2} = 1 + \epsilon/2 - \frac{\epsilon^2/4}{1 + \epsilon/2}$$

Substituting in Eq. (7) gives

$$x_2 = 1 + \epsilon/2 - \frac{\epsilon^2/8}{1 + \epsilon/2}$$

Hence, by the same reasoning used in Section 3.3

$$\left|\sqrt{1 + \epsilon} - x_1\right| < \left|x_2 - x_1\right| = \frac{1}{8}\left(\frac{1}{1 + \epsilon/2}\right)\epsilon^2$$

For $\epsilon \ll 1$, say $\epsilon = 0.1$, we can be sure that

$$\left|\frac{1}{8}\left(\frac{1}{1 + \epsilon/2}\right)\right| < 0.12$$

Thus we have shown that the magnitude of the difference between $\sqrt{1 + \epsilon}$ and $1 + \frac{\epsilon}{2}$ is less than a constant times ϵ^2. Therefore, the first order approximation for $\sqrt{1 + \epsilon}$ is

$$\sqrt{1 + \epsilon} \approx 1 + \epsilon/2 \tag{8}$$

Questions

1. Could you choose $x_0 = 1 + \epsilon$ as your starting point in generating the first order approximation for $\sqrt{1 + \epsilon}$?
2. Use the first order approximation to calculate $\sqrt{1.04}$. Give an upper bound for the error.
3. Use the relation $\sqrt{A + a} = \sqrt{A}\sqrt{1 + \frac{a}{A}}$ to calculate an approximate value of $\sqrt{50}$.
4. Use the relation $\sqrt{pq} = \sqrt{p} \times \sqrt{q}$ and the first order approximation to find $\sqrt{1.06 \times 10^4}$ and $\sqrt{2.6 \times 10^{-3}}$
5. A jet plane is 20 miles from the control tower of an airport and at an altitude of five miles.
   (a) Use the first order approximations for $\sqrt{A + a}$ to find the line-of-sight distance from the control tower to the plane.
   (b) What was the per cent error in your answer for (a)?

Chapter 4. SLIDE RULES

It is not at all tedious to add two 3-digit numbers, but multiplying them together is a chore. Although in most cases an exact answer is not required because the original quantities are themselves not known exactly, we often need a more exact answer than can be obtained by mental approximation. A slide rule fills this need admirably. It is nothing more than two pieces of wood, plastic, or metal with scales engraved on them, joined so that one slides on the other, but it can be used to multiply and divide quickly and with considerable accuracy. For example, a mental approximation applied to the following problem may yield

$$\frac{112 \times 17 \times 45 \times 87}{32 \times 43 \times 72} = \frac{112}{32} \times \frac{17}{43} \times \frac{45}{72} \times 87 \approx 4 \times \frac{1}{2} \times \frac{1}{2} \times 80 = 80$$

This calculation, worked out on a slide rule in about one minute, gave 75.1. The answer, worked out more precisely with a desk calculator, is 75.24.

4.1 Multiplication and Division of Powers of Two

In this section we shall put scales on a simple slide rule which will enable us to multiply powers of two together. Then, in later sections, we shall see how this slide rule can be made into one which can deal with any numbers.

The slide rule you need has unlabeled, equally-spaced lines on the back. There are two sets of 13 lines on the "fixed," outer part and two identical sets on the movable, inner part. To make reference to the different sets of lines or scales easier we shall arbitrarily give them names. We name the upper scale on the fixed part of the slide rule, the E scale, the upper edge of the sliding inner part, the F scale; its lower edge, the C scale. The lower fixed scale we shall name the D scale. Write the appropriate name at the extreme left end of each scale on your slide rule.

Label the center marks on the E and F scales with the number 0, the marks to the right of 0 on each part with increasing integers and to the left of 0 with decreasing integers (Fig. 4.1).

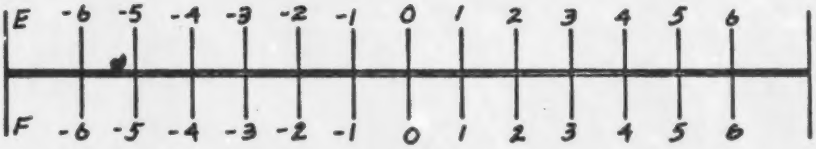

E
-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6
F
-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6

Fig. 4.1

Now suppose we wish to use the slide rule to add 3 + 2. In Fig. 4.2(a) the two scales are arranged so that 0 on the F scale coincides with 3 on the E scale. With this setting we can add a number to 3. To find 3 + 2, for example, we find 2 on the F scale and read the answer, 5, directly above on the E scale. Figure 4.2(b) shows the addition of a negative number and a positive one. If you look above -4 on the F scale you will find the answer to 3 + (-4). Note that without moving the E scale we can add to 3 any number between -6 and +3. In effect, what we have done in adding the two numbers is to add two displacements, the arrows in Fig. 4.2(a) and 4.2(b), to get a total displacement which is the sum we are seeking.

Fig. 4.2(a)

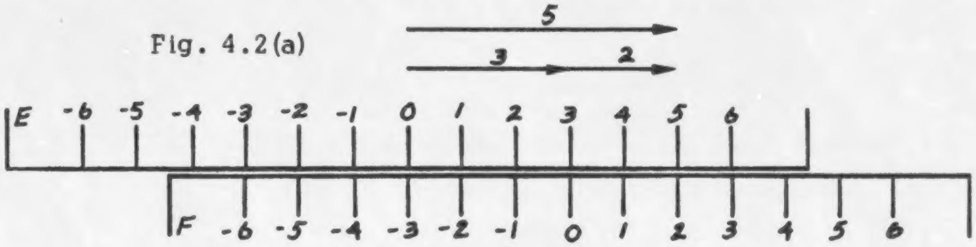
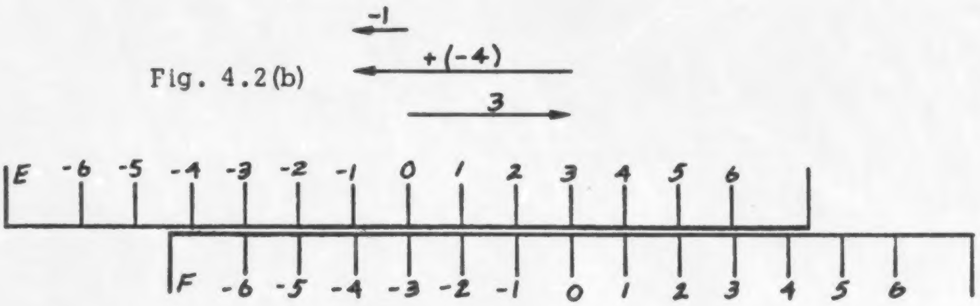

5
3
2
E
-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6
F
-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6

Fig. 4.2(b)


-1
+(-4)
3
E
-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6
F
-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6

Now consider scales C and D. We shall associate each integer on the E and F scales with the value of 2 raised to that integral power and label the C and D scales with these powers of two. This gives what is called a logarithmic scale. For example, we place the number 4 at the marks on the C and D scale directly below the 2 mark on the F scale (Fig. 4.3). Now each time you perform an addition, using the E and F scale, you are adding the exponents of 2 on the E and D scales. Thus you are performing a multiplication of the corresponding numbers on the C and D scales. To see why this is so, recall that $10^{m+n} = 10^m \times 10^n$ and just as with powers of ten, it is true that

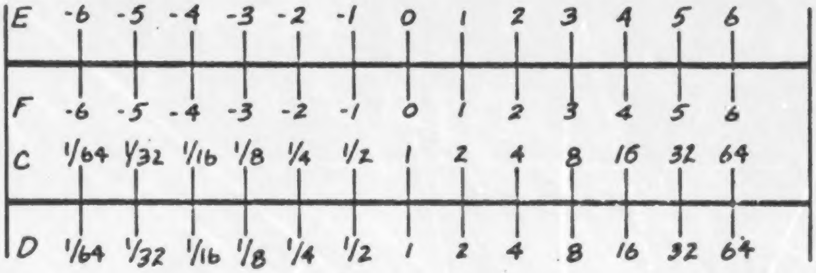

E -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6
F -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6
C 1/64 1/32 1/16 1/8 1/4 1/2 1 2 4 8 16 32 64
D 1/64 1/32 1/16 1/8 1/4 1/2 1 2 4 8 16 32 64

Fig. 4.3

$$2^{m+n} = 2^m \times 2^n$$

for all integers m and n, both positive and negative, and zero.

For example, 1 on the F scale coincides with the mark for 2 on the C scale and 3 on the E scale coincides with 8 on the D scale. Therefore, when we add 1 and 3 using the E and F scales to get 4, we are, in fact, adding the exponents of the numbers $2^1 = 2$ and $2^3 = 8$. This is equivalent on the C and D scales to multiplying 2 × 8.

$$\begin{aligned} 2^{1+3} &= 2^4 = 16 \\ &= 2^1 \times 2^3 = 2 \times 8 = 16 \end{aligned}$$

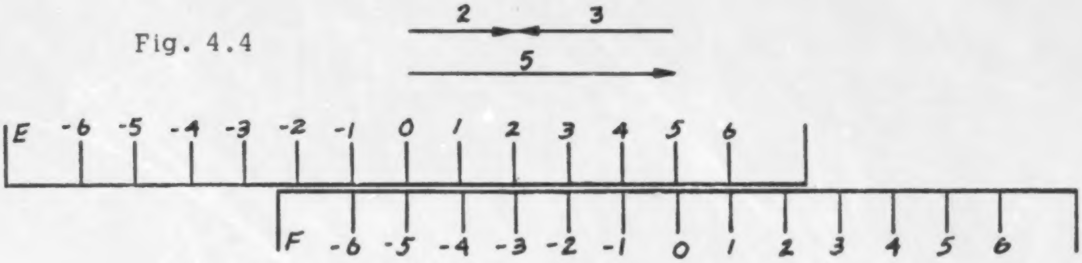
As you can see, 4 on the E scale coincides with 16 on the D scale.

When we add a negative number to a positive one, using the E and F scales as in Fig. 4.2(b), we are at the same time dividing one power of two by another on the C and D scales. This is true because if m and n are integers,

$$2^{m+(-n)} = 2^{m-n} = 2^m \times 2^{-n} = \frac{2^m}{2^n}$$

Thus in Fig. 4.2(b) we performed the division $\frac{2^3}{2^4} = 2^{-1} = \frac{1}{2}$. Any displacement to the left of a number is a subtraction and is equivalent to a division. For example, in Fig. 4.4 we have done the subtraction 5 - 3 = 2 using the E and F scales which is equivalent to $2^{5-3} = \frac{2^5}{2^3} = 4$.

Fig. 4.4



Questions

1. Draw rough diagrams showing the relative positions of the E and F scales on your slide rule after performing the following additions:

   (a) 1 + 4 (c) 2 + (-3)

   (b) -3 + (-1) (d) 0 + 2

2. Give the multiplication problem solved on your slide rule corresponding to each of the additions in Question 1. Write these multiplications both in exponential form and without exponents.

3. Write the following multiplications in the form $2^m \times 2^n$ where m and n are integers. Do each of the multiplications, using your slide rule. What addition is being performed in each case?

   (a) $8 \times \frac{1}{16}$ (c) 1×16

   (b) $\frac{1}{8} \times \frac{1}{4}$ (d) $\frac{1}{64} \times 64$

4. Use your slide rule to do the following divisions:

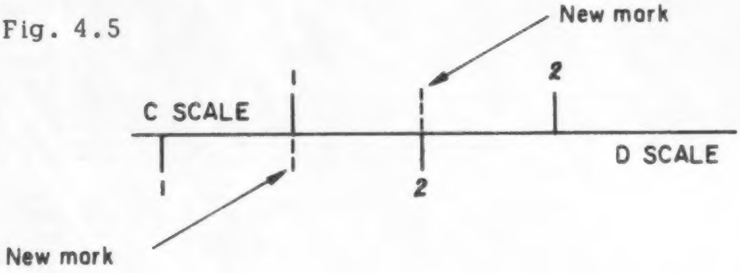
   (a) $\frac{64}{16}$ (c) $\frac{1}{2}$

   (b) $\frac{32}{4}$ (d) $\frac{1}{8}$

4.2 Non-integral Powers of Two

The numbers on the C and D scales you have labeled range from $\frac{1}{64}$ to 64. With these scales you can easily multiply any pair of these numbers (numbers which are integral powers of two) as long as the product is between $\frac{1}{64}$ and 64. But what if we wish to multiply and divide numbers that are not integral powers of two? It seems reasonable that numbers between integral powers of two on your slide rule can be represented by points between the ones already marked. But how are these intermediate marks to be determined? On a centimeter rule, marked off only in centimeters, if you wish to indicate where the half-centimeter marks should be placed you put marks halfway between the centimeter marks. We can do this, because the centimeter marks are equally spaced. On a slide rule, however, such is not the case. The numbers increase more and more rapidly for equal distances on the rule as one approaches the right-hand end. The interval on the left end corresponds to an increase of $\frac{1}{32} - \frac{1}{64} = \frac{1}{64}$, while an equal interval on the extreme right-hand side corresponds to an increase of 64 - 32 = 32. What number does the point halfway between 1 and 2 on your slide rule correspond to? Mark off the point halfway between 1 and 2 on both the C and D scales. Now set the 1 on the C scale at this halfway mark. If we multiply this unknown number by itself as shown in Fig. 4.5, the result is 2.

Fig. 4.5



The number which when multiplied by itself yields 2 is $\sqrt{2}$. This is the unknown number we are looking for. The square root of 2 is close to 1.41, so label this point as 1.41 on the C and D scales.

Now that we know that the point halfway between 1 and 2 on the C and D scales should be labeled 1.41, we can find the values of the points halfway between the other markings. For example, if we multiply 4 by 1.41 using the slide rule, we find that the answer is at the point halfway between 4 and 8. But we also know that 4 × 1.41 = 5.64 so this point should be labeled 5.64.

We have found the value of the mark halfway between 1 and 2 on the C and D scales to be $\sqrt{2}$. What about the corresponding mark on the E and F scales? This corresponding mark lies halfway between 0 and 1 on these scales and since the numbers on the E and F scales increase uniformly, the midpoint has the value $\frac{1}{2}$. We have labeled our slide rule so that the numbers on the E and F scales are the powers to which 2 must be raised to get the values of the corresponding points on the C and D scales. Since thus far we have studied only integral powers, we have written something new, namely $\sqrt{2} = 2^{1/2}$. That this is reasonable is borne out by the fact that we can use $a^n \times a^m = a^{n+m}$ to get $2^{1/2} \times 2^{1/2} = 2^{1/2 + 1/2} = 2^1 = 2$. Just as we can continue to find values of points on the C and D scales, we can extend our ideas about fractional powers of two to many fractions by considering points halfway between known values. This will be seen in the following questions.

Questions

1. (a) Find the value on the C and D scales of each point halfway between the original marks on the slide rule.
   (b) To what number on the E and F scales does each correspond?
2. Use the method described in Section 3.6 for approximating square roots to find the number halfway between 1 and 1.41 on your C and D scales.
3. Now that you know the value of the point halfway between 1 and 1.41 (Question 2), multiply it by other known values on your slide rule to find the values of some other unknown points.

4. How many points could be labeled, using the half values on your slide rule and the answer to Question 3?

5. (a) How can the cube root of 2 be written in terms of fractional exponents?

   (b) How is each of the following obtainable by taking square and cube roots

$$2^{1/4},\ 2^{1/3},\ 2^{1/6},\ 2^{1/8},\ 2^{1/12}$$

4.3 A Power-of-Ten Slide Rule

As you have just found out, the apparent limitation of our slide rule of being able to treat only those numbers which are integral powers of two can be overcome. Another limitation is that it can handle only multiplications and divisions between $\frac{1}{64}$ and 64. By making a sufficiently long slide rule we can deal with numbers as large as or as small as we wish, at least in principle, but since the slide rule is supposed to be convenient and easy to use, this would defeat the whole purpose of the instrument.

The solution to this problem lies in the fact that any multiplication or division can be divided into two parts, one involving numbers between 1 and 10 and the other involving only powers of ten. For example, $(1.65 \times 10^6) \times (1.21 \times 10^2) = (1.65 \times 1.21) \times 10^8$. Thus we need only multiply and divide numbers which are between 1 and 10. It turns out, therefore, that we need only that segment of the slide rule containing the numbers between 1 and 10. The rest is superfluous.

It is not clear that this is enough. If we have a slide rule which includes only the numbers from 1 to 10 and try to multiply, say, 6 × 6 by the method we have described, then the answer will not appear on the slide rule; it would lie beyond the end of the rule. Similarly, if we try to divide 2 by 9 the answer will not appear. However, a slide rule including the numbers from 0.1 to 100 will take care of such contingencies. This is because the product of any two numbers between 1 and 10 is less than 100 and the quotient of any two numbers between 1 and 10 is greater than 0.1. In fact, as you will see later, we can eliminate the need for this extended range of numbers, but first, we shall construct a slide rule covering three decades, from 0.1 to 100, from 10^{-1} to 10^2, to see how we can use a slide rule with a range of 1 to 10 to handle any numbers.

Erase all the numbers you have put on your slide rule. The two new scales you will construct will also be called C and D. Label the first mark at the left on both the C and D scales with the number 0.1. Label the fourth mark to the right 1. Thus, adding the distance between the first and fourth mark will correspond to multiplying 0.1 by 10. The fourth mark to the right of 1 should therefore be labeled 10 and the twelfth mark labeled 100. The resulting C and D scales are shown in Fig. 4.6.

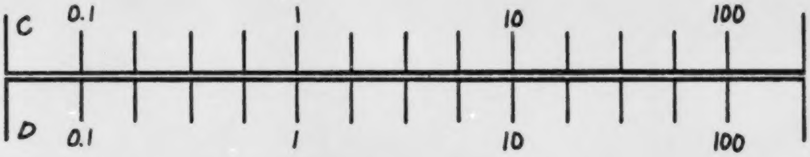


Fig. 4.6

Questions

1. (a) If we multiply the value of the point halfway between 1 and 10 by itself we get 10. Use this fact to find the value at this point.

   (b) Use the answer to (a) to find the values corresponding to the points halfway between the ends of the other two decades.

   (c) Find the value of the point one-quarter of the way between 1 and 10. Use this value to find values corresponding to all the rest of the marks on the C and D scales.

4.4 Division and Multiplication Using Only a One-Decade Slide Rule

With the new C and D scales we have marked off, we can divide and multiply any pair of numbers between 1 and 10. Now we shall use this set of scales to show that in fact one can do the same thing using only the middle portion of a three-decade slide rule.

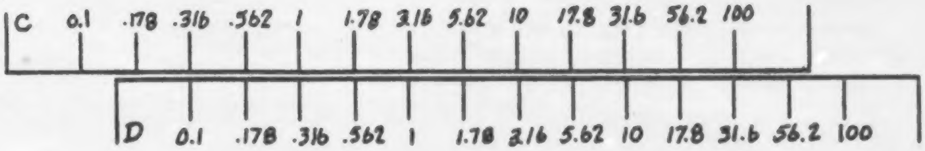

C
0.1
.178
.316
.562
1
1.78
3.16
5.62
10
17.8
31.6
56.2
100
D
0.1
.178
.316
.562
1
1.78
3.16
5.62
10
17.8
31.6
56.2
100

Fig. 4.7

First, let us see how we can perform divisions using only the middle decade. The division $\frac{1.78}{5.62}$ is indicated on the slide rule shown in Fig. 4.7. First, 1.78 is located on the D scale and the 5.62 on the C scale positioned directly over it to enable one to go back by the length corresponding to the number 5.62. The answer, close to 0.316, can be read directly under the 1 on the C scale. However, notice that if we multiply the answer 0.316 by 10 (we do this by reading the number on the D scale directly under 10 on the C scale) we see from Fig. 4.7 that the 10 on the C scale is almost directly over 3.16 on the D scale. What about other divisions? Clearly, either the 1 or the 10 of the C scale must be over the central decade of the D scale in any division involving two numbers between 1 and 10. If the 1 is over this portion of the D scale, the correct answer can be read under it without further ado. If not, then the 10 of the D scale is over a number which is ten times the desired answer. The fact that this is ten times too large is unimportant, because we can easily find the location of the decimal point by estimation. It is clear, for example, that the answer to the division described above, $\frac{1.78}{5.62}$ lies somewhere between 0.1 and 1.

In multiplication problems, as in division problems involving two numbers between 1 and 10, the answer does not always fall within the 1 to 10 decade. But again, as in division, there is a simple way to get the answer. To multiply 3.16 by 5.62 we set the C and D scales as shown in Fig. 4.8(a). The answer falls beyond the end of the center decade of the D scale.

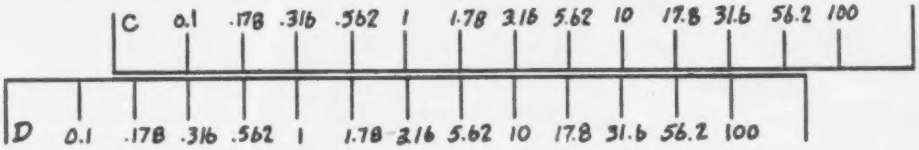

C
0.1
.178
.316
.562
1
1.78
3.16
5.62
10
17.8
31.6
56.2
100
D
0.1
.178
.316
.562
1
1.78
3.16
5.62
10
17.8
31.6
56.2
100

Fig. 4.8(a)

As you can see from the figure, it is close to 17.8. However, if we start over again and divide 3.16 by 10 as shown in Fig. 4.8(b), you can see that the factor 5.62 falls directly over 1.78, which is just one-tenth of the answer. Again, we are not concerned about the decimal point because we always find it by estimation. The important thing is that by reversing the end of the center decade of the C scale that we place over one of the factors, we can find in the center decade the correct digits of the answer. Thus we can do any multiplication of numbers between 1 and 10 using only the parts of the C and D scales between 1 and 10. First we try the usual procedure for multiplication. If the second number, on the C scale, is not over the center decade of the D scale, we move the 10 of the C scale over the first number and then the digits of the answer will certainly appear beneath the second factor in the multiplication. We can, therefore, dispense with the other two decades.

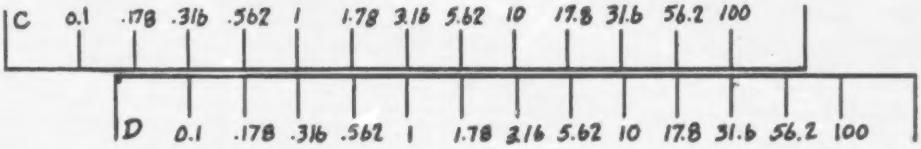


Fig. 4.8(b)

Questions

1. Which of the following division problems would have answers lying under the 1 of the C scale and which would have answers beneath 10?

   (a) $\frac{5.62}{3.16}$

   (b) $\frac{1.78}{5.62}$

   (c) $\frac{1.00}{1.78}$

2. Perform the following divisions, using the slide rule only to find the digits. Use only the center decade of the C and D scales. Find the correct placement of the decimal point by estimation.

   (a) $\frac{0.0178}{5.62}$ (b) $\frac{5620}{0.178}$

3. Perform the following multiplications, using only the center decade of your slide rule.
   (a) 0.00178×0.0178
   (b) 3.16×56.2
   (c) 17.8×0.0562

4.5 Commercial Slide Rule Scales

Not every division on a ruler is marked with a number. Usually the number marks correspond to integral numbers of inches or centimeters. The subdividing marks, being equally spaced, have values that can easily be determined by inspection and need not be labeled. In the interval between 0 and 1 cm on a centimeter scale there are ten subdividing marks, each mark corresponding to 0.1 cm. The numbers you placed on your power-of-ten slide rule you found by taking successive square roots of 10 and are not succesive whole numbers, and do not make a decimal scale; it is therefore awkward to use. To locate the points on a power-of-ten slide-rule scale corresponding to any numbers we first make a table (Table 4.1) of the displacements* and the corresponding numbers using the information on the scales from 1 to 10 in Fig. 4.7.

TABLE 4.1

Displacement	Number
0	1.00
2.5	1.78
5.0	3.16
7.5	5.62
10.0	10.00

---

*In the table, a displacement of 2.5 units equals 1 cm.

Then we make a graph using the data in this table, drawing a smooth curve exactly connecting all the points as shown in Fig. 4.9. (For better accuracy we could calculate intermediate points to add to the information in Table 4.1 and make a larger graph than that in Fig. 4.9.) We then use the graph in Fig. 4.9 to read off the displacements for the numbers we wish to put on the slide rule and make a second table (Table 4.2 is an abbreviated form of such a table) which we can use to make a power-of-ten slide rule having convenient numbers and subdivisions.

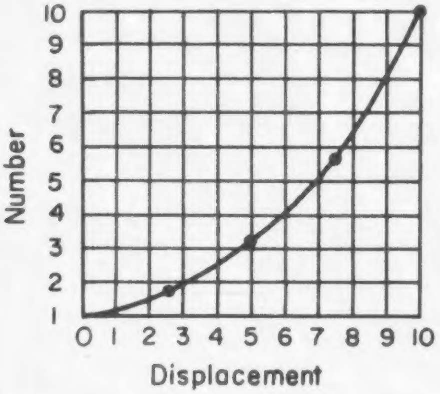


Fig. 4.9

TABLE 4.2

Number	Displacement
1	0
2	3.01
3	4.77
4	6.02
5	6.99
6	7.78
7	8.45
8	9.03
9	9.54
10	10.00

On your slide rule you will find an "L" scale (used for finding logarithms of numbers) marked off with equal divisions. This scale goes from 0 to 10 and is the same length as the C and D scales so you can use it to measure displacements and check the values in Table 4.2.

A slide rule made commercially has points marked on it which correspond to convenient numbers. Look at the engraved points on the C and D scales on the other side of the slide rule with which you have been working corresponding to the integers labeled 1, 2, 3, etc. Each of the intervals between these numbers is subdivided. However, these intervals do not have the same number of subdivisions. The interval between 1 and 2 is divided into ten labeled parts, 1.1, 1.2, 1.3, ... 1.9, and each of these in turn is divided into ten parts by unlabeled marks so the smallest divisions correspond to 0.01. The space between 2 and 4 is also divided into ten parts, but since there is less space each of these ten is divided into only five parts. Thus the smallest subdivision in this range corresponds to 0.02. Between 4 and 10 the intervals between integers are divided into 10 large intervals, but the distance between integers is so short that each of these intervals is divided into only two small intervals, each equal to 0.05. As you can see, one must be careful in reading the scales on a commercial slide rule.

Questions

1. Use the graph in Fig. 4.9 to find
   (a) the number on a slide rule corresponding to a displacement of 3.5.
   (b) the displacement corresponding to the number 6.5 on the D scale.
2. Perform the following multiplications on a commercial slide rule. Use exponential notation in locating the position of the decimal point.
   (a) 31.7×45.6 (c) 863×749
   (b) 0.37×7.44 (d) 0.000845×0.000079
3. Perform the following divisions on a commercial slide rule:
   (a) $\frac{0.00000049}{43}$ (c) $\frac{362}{0.0043}$
   (b) $\frac{4.3 \times 10^{11}}{376}$ (d) $\frac{1.07}{4070}$

4.6 Multiple Multiplication and Division

The slide rule is ideal for long series of calculations. The sliding crosshair can be set to the result of intermediate calculations to keep track of them, but there is no need to read the answer for each multiplication or division. For example, consider the product $22 \times 2.3 \times 8.9 \times 4.8$.* First, you multiply 22 by 2.3 starting with the left end of the C scale over 22, setting the sliding crosshair over the answer on the D scale. Then, without bothering to read the answer, set the right-hand end of the C scale so that it coincides with the crosshair. You are now ready to multiply the product 22×2.3 by the next factor, 8.9. To do this you simply move the crosshair to 8.9 on the C scale. The answer lies directly below on the D scale, but you do not bother to read it; you just move the right-hand end of the C scale to this point and then move the crosshair to 4.8 on the C scale to complete the calculation. Now the answer can be read from the position of the crosshair on the D scale. The digits are 216.

To find the decimal point you make a simple approximation:

$$22 \times 2.3 \times 8.9 \times 4.8 \approx 20 \times 2 \times 9 \times 5 = 1800$$

Thus the correct answer is 2160.

A series of divisions is even easier to do. Take, for example, the calculation of $\frac{1}{2.2 \times 4.8 \times 5.2}$. To find the answer quickly and easily without bothering about intermediate answers, you first move 2.2 on the C scale over the left end of the D scale to divide $\frac{1}{2.2}$. Placing the crosshair over the answer at the end of the C scale, you can now divide by 4.8 by moving the C scale so that 4.8 on this scale coincides with the crosshair. Next the crosshair is moved to the answer under the end of the C scale. The final division by 5.2 can now be made by moving 5.2 on the C scale to coincide with the crosshair. The final answer is then read on the D scale below the end of the C scale. The digits in the final answer are 182. Making a rough approximation of the problem we get

---

*Follow each step in the examples in this section with your own commercial slide rule.

$$\frac{1}{2 \times 5 \times 5} = \frac{1}{50} = 0.02$$

so the correct answer is 0.0182.

The tricks discussed in the two examples above are particularly useful in solving calculations that are a combination of both multiplication and division. Suppose you have to calculate

$$\frac{3 \times 7 \times 2.5}{5 \times 4 \times 1.9}$$

The easiest way to do the calculation is to divide 3 by 5, multiply the result by 7, then divide by 4, multiply by 2.5 and finally divide by 1.9 without reading any answer except the final one to get the digits 138. Approximation places the decimal point and the correct answer is 1.38.

A vast amount of arithmetical drudgery can be saved by using a slide rule to perform multiplications and divisions and the results are accurate enough for nearly all purposes. Once you have learned how to do different kinds of calculations, the only source of error is in reading the scales. After you have had sufficient practice in reading the scales, you will find that you can calculate very rapidly with a slide rule and make very few errors.

Questions

Perform the following calculations without reading any of the intermediate products.

1. (a) $14 \times 2.5 \times 13 \times 13$
   (b) $1.55 \times 2.37 \times 110 \times 226$
   (c) $7.8 \times 197 \times 2.00 \times 7.13$
   (d) $11.7 \times 9.83 \times 10^{-6} \times 3.05 \times 10^{-8}$

2. (a) $\frac{1}{2.3} \times \frac{1}{9.8} \times \frac{1}{127}$

   (b) $\frac{1}{3} \times \frac{1}{4.006} \times \frac{3}{7.1}$

   (c) $\frac{1}{6.1 \times 10^4} \times \frac{1}{5.2 \times 10^{-3}}$

   (d) $\frac{1}{3.06 \times 10^4 \times 2.14 \times 10^3}$

3. (a) $\frac{37.6 \times 12.4 \times 8.3}{2.7 \times 3.78 \times 4.11}$

   (b) $\frac{63.4 \times 4.73 \times 7.79}{21.2 \times 2.86}$

   (c) $\frac{8.72 \times 10^3 \times 3.64 \times 10^{-7} \times 11.2 \times 10^4}{11.1 \times 10^6 \times 2.34 \times 6.38 \times 10^{-3}}$

   (d) $\frac{0.0037 \times 6.5 \times 10^{10} \times 873}{41.3 \times 18 \times \frac{1}{127} \times 8.81}$

   (e) $\frac{2.718 \times 3.00 \times 10^8}{3.14 \times \frac{1}{127} \times 9.80 \times 0.667 \times 4}$

4. On the C and D scales of a 10-inch slide rule, what is the relative uncertainty in reading a number between (a) 1 and 2? (b) 3 and 4? (c) 9 and 10?

4.7 Constant Factors, Ratios, and Uncertainty

Many times in making calculations we encounter situations in which we have to multiply a series of numbers by the same constant factor. For example, in making a map we have to multiply a large number of measured distances by a scaling factor to get the correct lengths to put on the map. This is easy with a slide rule. All we have to do is set the end of the C scale once (or at most twice) directly over the constant scaling factor and then just move the crosshair to perform each successive multiplication.

For example, if we set the 1 on the C scale over 3 on the D scale we can multiply 3 times any number from 1 to 3.33 merely by moving the crosshair to the number on the C scale by which we wish to multiply by 3 (Fig. 4.10), and then reading the answer on the D scale. For numbers greater than 3.33 we simply set the other end of the C scale over 3 on the D scale.

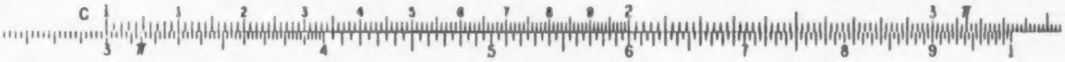


Fig. 4.10

Similarly, if we set the 7 on the C scale over 3 on the D scale, as shown in Fig. 4.11, the ratio of any number (from 1 to 4.28) on the D scale to the number directly above it on the C scale is $\frac{3}{7} = 0.428$. For numbers on the D scale between 0.428 and 10 we set the left end of the C scale over 4.28 on the D scale.

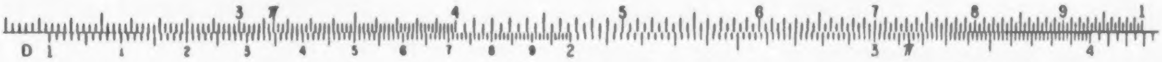


Fig. 4.11

The uncertainty in reading any scale is a fixed small distance along the scale. For example, one might be able to read a centimeter scale to within 0.02 cm. This uncertainty in a length reading matters much more for short lengths than for long ones when we are concerned with relative uncertainty. Consider two extreme cases: a length of 0.50 cm with an uncertainty of 0.02 cm has a relative uncertainty of $\frac{0.02 \text{ cm}}{0.50 \text{ cm}} \times 100 = 4$ per cent; a reading of 20 cm with the same uncertainty of 0.02 cm has a relative uncertainty of $\frac{0.02 \text{ cm}}{20 \text{ cm}} \times 100 = 0.1$ per cent.

On a "10-inch" slide rule, the C and D scales are about 25 cm long and, reading from the left end, 1 cm corresponds very nearly to a factor of 1.1. Since the divisions on the scale between 1.0 and 1.1 are almost equal, 0.02 cm represents a factor close to 1.002. Suppose you move the left end of the C scale to any position along the D scale. An uncertainty of 0.02 cm in the reading on the D scale still corresponds to a factor of 1.002. Thus the fractional uncertainty in reading a slide rule is constant, and all readings on the C and D scale have an uncertainty of about 0.2 per cent.

Questions

1. To what multiplication factor does a distance of 1 cm on the C and D scales correspond?
2. If you move the 1 on the C scale to a point directly above 1.50 on the D scale,

   (a) what is the ratio of any number on the C scale to the one directly below it?

   (b) what is the ratio of any number on the D scale to the number directly above it?
3. On a commercial slide rule there are two adjacent scales labeled A and B. Each of these is a two-decade scale and the decades are just half as long as the C and D scales. What is the relation between a number on the A scale and the number directly below it on the D scale? Can you explain why there is this relation?
4. Problem with student?

Chapter 5. AUTOMATIC COMPUTATION

In the previous chapter you have learned a number of techniques for calculating effectively. We now turn to the problem of calculating effectively when the task involves repetition in one form or another. For this purpose it is often convenient to use a computer.

Our motivation for "programming" a computer (writing instructions that tell the computer how to carry out a calculation) is similar to the motivation for building a machine to mass-produce a product: the time and money required to build a machine to stamp out "widgets" is greater than the cost of making one widget by hand; but after the initial investment, widgets can be produced cheaply in quantity. Once a program has been prepared, it is easy to have it executed many times by a computer. Although computers can calculate many times faster than the human brain, speed alone is not the essence of the power of computers. No matter how fast a computer can calculate, doing a one-shot job on a computer is a waste of time if it is easier to punch keys on an electronic calculator (or even do pencil and paper calculations) than to write a program to get a computer to do it. Hence, a single calculation, however involved, seldom requires the use of a computer program if it is to be used only once.

There is a further benefit that derives from learning how to program a computer. A computer has a small "vocabulary" and cannot make the subtle judgments of the meanings of words and symbols that human beings are capable of. Therefore, to write a program for a computer, one must learn to think carefully in order to give the precise instructions to the computer that it needs in order to carry out the desired calculations.

5.1 Programs

Suppose you are asking another person to average five numbers, using a desk calculator. The request "Please average these five numbers" will suffice if the other person is knowledgeable in mathematics and competent in the operation of the desk calculator. "Add up these five numbers and divide by five" is a bit more explicit. However, suppose one is dealing with a very inexperienced helper who is going to use a certain desk calculator to find the average. If the computation is a one-shot job, it would be easiest to do it oneself; however, suppose it is to be carried out a great many times. One might then have to spell out this task in detail as follows:

1. Press the "clear" button*
2. Punch the first number in the keyboard and press the "+" button
3. " " second " " " " " " " " "
4. " " third " " " " " " " " "
5. " " fourth " " " " " " " " "
6. " " fifth " " " " " " " " "
7. Punch 5 in the keyboard and press the "÷" button
8. Record on paper the number displayed.

Such a set of instructions is called a program. This very simple program has many of the features typical of programs for mathematical calculations, including:

(a) Numbers are entered. This is referred to as input.

(b) Computations are performed and intermediate results stored.

(c) Results are recorded. This is referred to as output (in the above example the output consists of a single number).

(d) The instructions are to be carried out in order, starting at the top. (At the end of each step the affix "and proceed to the next step" is implicit.)

(e) The program can be applied not only to one specific set of input numbers, but to arbitrary sets; therefore, it may be repeatedly useful.

---

*This erases from the computer any numbers it is storing as a result of carrying out a previous program.

Note that this program, though more specific than the original statement "average these five numbers," still has meaning only in a specific context, involving a given type of desk calculator. It is necessary to understand the context before a program is completely intelligible.

Though we usually do not refer to them as such, the recipes in cookbooks are, in fact, programs. There the context assumed is a properly equipped kitchen, plus a cook familiar with the elementary techniques and vocabulary of cooking. Similarly, the instructions one might give a stranger for getting to one's house are, in effect, a program. One usually assumes then a driver who can count traffic lights, recognize landmarks, etc.

In computer programs the context which is assumed involves such things as memory storage locations, conventions about how storage locations are named, and how numbers are entered into them and retrieved from them, conventions as to what arithmetic operations are available, how the input and output of numbers can be handled, etc. Rather than listing all of these conventions at the outset, we will let them emerge as we proceed.

Let us re-express our program to average five numbers in language that refers less specifically to a desk calculator. We need the idea of a device in which a number can be stored. The common name for such a device is storage register, or simply register. Here we will need two registers, which we will name X and S. Register X will correspond to the keyboard of the desk calculator, and register S to the "display."

Generally, in computers, a number can be retrieved from a register, with the number stored remaining intact in the register (this is called "non-destructive read-out"). When a number is read into a register the number previously stored is, of course, lost.

Using the storage registers X and S our program which we will refer to as Program 1 might be as follows:

PROGRAM 1

1. Store 0 in S.
2. Read the next input number and store it in X.*
3. Compute X + S and store the result in S.
4. Read the next input number and store it in X.
5. Compute X + S and store the result in S.
6. Read the next input number and store it in X.
7. Compute X + S and store the result in S.
8. Read the next input number and store it in X.
9. Compute X + S and store the result in S.
10. Read the next input number and store it in X.
11. Compute X + S and store the result in S.
12. Compute S/5 and store the result in S.
13. Write S.

Note that at the end of each step all numbers are left in registers. This is fundamental in computer programming; numbers can never be left in limbo, and it would be incorrect to replace steps 11 and 12 by

11. Compute X + S
12. Divide the result of step 11 by 5.

Obviously a statement such as "compute X + S" must mean "compute the contents of X plus the contents of S." For the sake of brevity, we prefer not to incessantly include the words "contents of." Thus a symbol such as X does double duty, serving both as the name of a storage register and as a symbol for the contents of that register. Which meaning is intended is fortunately almost always clear from the context.

Now assume, for example, that Program 1 is executed using the input data

20
10
45
15
60

---

*As we will see shortly when we discuss input number conventions, in this program the first input number is read in on this step.

Such a list of input numbers is always entered into the computer in order, starting at the top, as input is called for by the program. Therefore in step 2 "the next" input number is the first, namely 20, on step 4 "the next" is 10, etc.

Table 5.1 shows the contents of registers X and S after each step.

TABLE 5.1

Step	Contents of X	Contents of S
1	?	0
2	20	0
3	20	20
4	10	20
5	10	30
6	45	30
7	45	75
8	15	75
9	15	90
10	60	90
11	60	150
12	60	30
13	60	30

The final answer, written out on step 13, is, of course, 30. Any horizontal line in Table 5.1 gives a "snapshot" of the numbers stored at the corresponding intermediate point in the computation. Such a record of the history of the execution of a program with specific input data is called a trace.

Note very carefully that whereas S is always the same storage register, its contents (also referred to as S in Program 1), changes during the calulation, just as the reading of the desk calculator display changes. The distinction between a register and its contents, and the fact that the value of the latter depends on what point has been reached in the program, must always be clear when one is dealing with programs.

Questions

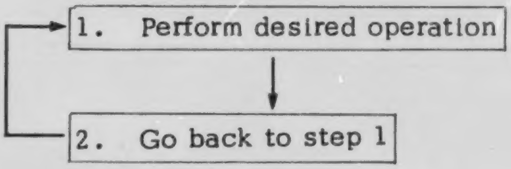
1. Do a trace for Program 1 applied to the following input data:
    (a) 14, 7, 3, -1, 8
    (b) 25, 10, 0, 6, -7
2. Write programs similar to Program 1 which do each of the following:
    (a) Compute the average of four numbers.
    (b) Compute the average of six numbers.
    (c) Compute the product of five numbers.
    (d) Compute the sum and the sum of the squares of five numbers.
3. Write an (English language) program for changing a flat tire. What context are you assuming?
4. What explanation can you give for the question mark on step 1 in Table 5.1?

5.2 Loops and Branches; Flow Charts

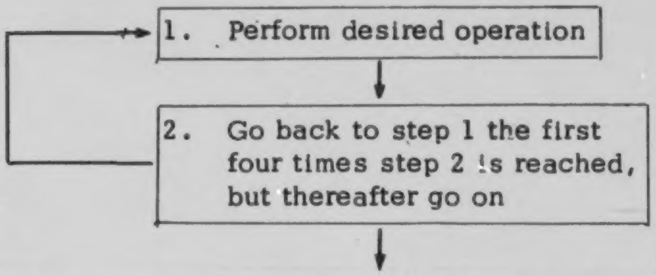
An obvious inefficiency of Program 1 is that the same pair of steps is repeated five times. If we modified the program to average, say, 100 numbers, this inefficiency would become painful indeed.

Of course, in dealing with a human helper we could say something like "repeat thus-and-such steps until all input numbers have been taken care of." However, such a statement is not sufficiently explicit when one is dealing with an automatic computer. How can we make a program in which a certain portion is repeated many times?

If step 1 is to perform a desired operation and the instruction step 2 is to go back to step 1, the operation will be performed many times, but we have made no provision for determining how many times. This is an example of an infinite loop, obviously to be avoided in practice. Such a program is frequently diagrammed as a flow chart, in which arrows indicate the "flow of control":



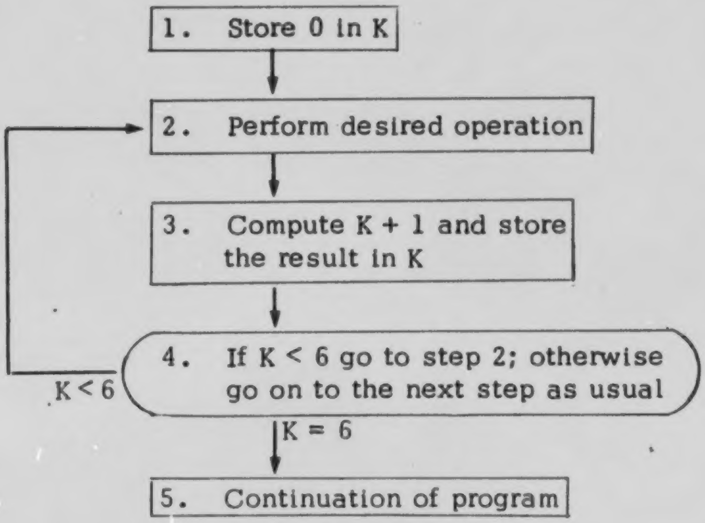
As our next attempt we might try:


1. Perform desired operation
2. Go back to step 1 the first four times step 2 is reached, but thereafter go on

This is a correct program, and could correctly guide a human capable of doing the counting called for in step 2. However, computers unaided by programs cannot count. The program must include some explicit device for counting.

Therefore we introduce another storage register, which we arbitrarily name K, in which to store a count of the number of times the "desired operation" has been executed. This adds a good deal of complexity to the logical structure of the program, which now appears as follows in Program 2:

PROGRAM 2


1. Store 0 in K
2. Perform desired operation
3. Compute K + 1 and store the result in K
4. If K < 6 go to step 2; otherwise go on to the next step as usual
K < 6
K = 6
5. Continuation of program

The first time step 3 is reached the contents of K are changed from 0 to 1, following which K stores the number 1, corresponding to the fact that the "desired operation" has been performed once. On the second pass through step 3, the contents of K are incremented to 2, etc., so that each time step 3 is completed K stores the number of times that step 2 has been executed. Thus register K does function correctly as a "counter."

Step 4 is a branch: two paths of control lead out of step 4 on the flow chart. (We have adopted the convention of drawing oval-shaped blocks around such branch points.) The "If $K < 6$" test in step 4 is a test of which computers are capable, and this "*If statement*" operation is very fundamental in computer programming. Between steps 2 and 4 we have a loop, which is cycled through five times during the running of the program.

Step 1, which sets the counter to zero, is essential; without it the contents of K, required on the first pass through step 3, would be undefined. The operation of step 1 is called *initialization*, which means the setting up of initial values in registers used in later computations.

We have in Program 2 the essence of the most important way in which computer programs take advantage of repetitive features of calculations. The point is that the instructions for the operations in step 2 need be written only once, even though they are performed many times.

Now let us go back to our original Program 1 which averages five numbers. Reorganizing it into the form of Program 2, we obtain Program 3.

PROGRAM 3

```
1.  Store 0 in S.
2.  Store 0 in K.
3.  Read an input number and store it in X.
4.  Compute $X + S$ and store the result in S.
5.  Compute $K + 1$ and store the result in K.
6.  If $K < 5$  go to step 3.
7.  Compute $S/5$ and store the result in S.
8.  Write S.
9.  Stop
```

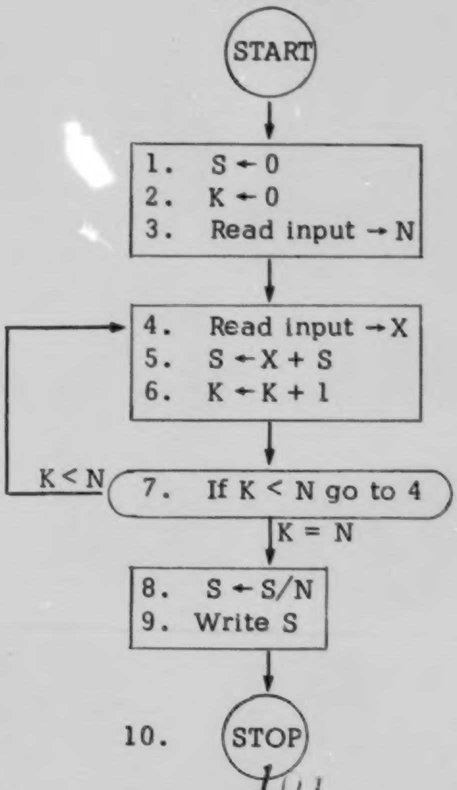
Let us abbreviate such statements as "store 0 in S" as "$S \leftarrow 0$." Our program then can be written in a briefer form (which incidentally is quite close to a program written in the BASIC or the FORTRAN computer language).

PROGRAM 3 (Abbreviated Notation)

```
1.  $S \leftarrow 0$
2.  $K \leftarrow 0$
3.  Read input $\rightarrow X$
4.  $S \leftarrow X + S$
5.  $K \leftarrow K + 1$
6.  If $K < 5$ go to line 3
7.  $S \leftarrow S/5$
8.  Write S
9.  Stop
```

A program that computes the mean of precisely five numbers is not of much general usefulness. However, we can easily generalize our program so as to calculate the mean of an arbitrary number N of values. Let us assume that the input consists of the value of N followed by the N values to be averaged. We shall need an additional register to store N; in fact we may as well call this new register by the name N, as our aid in remembering what it is used for. This time we give the program (Program 4) in the form of a flow chart.

PROGRAM 4


START
1. S ← 0
2. K ← 0
3. Read input → N
4. Read input → X
5. S ← X + S
6. K ← K + 1
K < N
7. If K < N go to 4
K = N
8. S ← S/N
9. Write S
10.
STOP

Note that only a small part of the program embodies the mathematical operations of averaging — in fact, only lines 5 and 8. (We will henceforth refer to the steps of programs as lines.) The business conducted in the rest of the program is referred to by the picturesque name of housekeeping — initializing, counting, getting input data into the right places, etc. This is essential in computer programs because, like the very inexperienced helper, computers don't know enough to do any of this without being told. Often there are many different ways of organizing housekeeping operations but no matter what way is used considerable ingenuity is required to keep the housekeeping free of "bugs," just as in real life.*

Questions

1. Which lines of Program 4 are examples of each of the following:
   (a) A loop
   (b) A branch
   (c) Initialization.
2. In Program 4
   (a) How many times is the loop traversed?
   (b) How many times is the "K < N" path, returning from line 7 to line 4, traversed?
3. What is the result of applying Program 4 to the following sets of input data:
   (a) 5, 6.1, 6.6, 6.3, 6.4, 6.1
   (b) 7, 1, 2, 3, 4, 5, 6, 7
   (c) 4, 1.3, 2.0, 3.1, 0.4, 5.1, 7.6, -1.2
   (d) 8, 1.1, 2.3, 4.6, 5.1, 6.2
   (e) -2.5, 6.1, 1.5, 8.3, 9.11

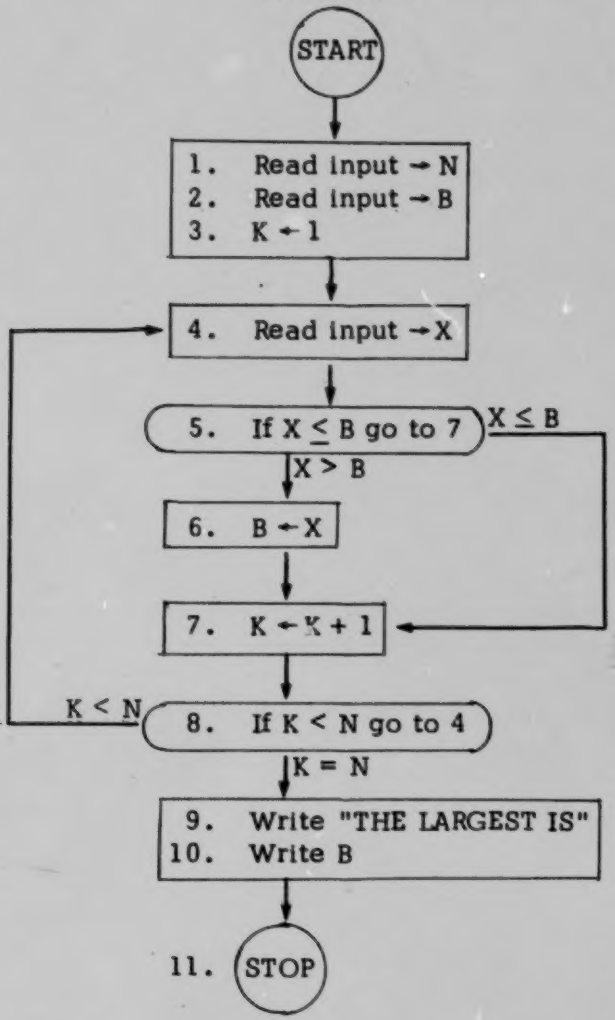
---

*In computer and electronic jargon "bugs" are errors in writing a program (or wiring a circuit) and "debugging" is the process whereby they are located and corrected.

4. In our development of programming, the choice of register names is to a large extent arbitrary, i.e. a matter of free choice for the programmer. (Often names with mnemonic significance are chosen, as "N" in Program 4, and also "S" for "Sum." This, however, is optional.) To illustrate this, write an alternate version of Program 4, in which the names "Q5," "J2," "J9," and "A7" are used in place of "S," "K," "N," and "X."
5. Let the input data to Program 4 be 3, 1.2, 2.6, 3.4. Do a trace as in Table 5.1, showing the history of the contents of registers K, X, and S.
6. Modify Program 4 so that it will compute the variance of the input data as well as the mean. The variance is the average of the squares of the values, minus the square of the mean. (Use another register, named S2, for the sum of the squares of the values read on line 4.)
7. Do a trace of the program you wrote in answer to Question 6 using as input data 3, 2, 3, 5. Does the result convince you that your program is free of bugs?

8. Program 5 is designed to find the largest of a set of N numbers:

PROGRAM 5


START
1. Read input → N
2. Read input → B
3. K ← 1
4. Read input → X
5. If X ≤ B go to 7
X ≤ B
X > B
6. B ← X
7. K ← K + 1
K < N
8. If K < N go to 4
K = N
9. Write "THE LARGEST IS"
10. Write B
11.
STOP

(a) What is assumed about input data?

(b) Do a trace for the input data 3, 5, 4, 6.

(c) How many branches does this program have?

(d) Explain the "bypass" from line 5 down to line 7.

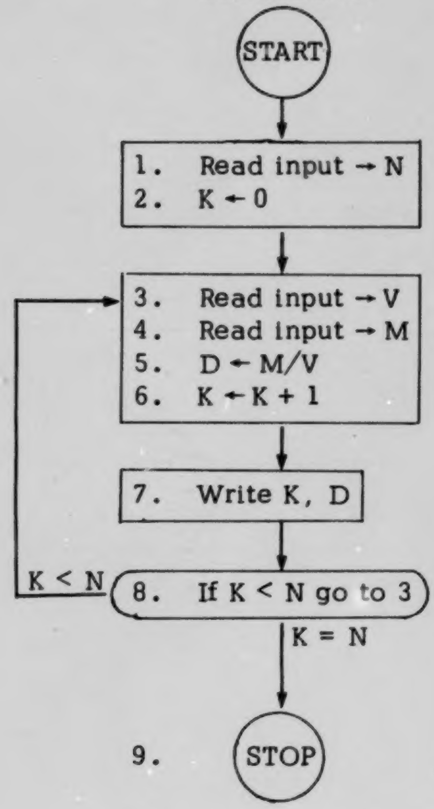
(e) How many times is the return path from line 8 back to line 4 traversed? (Let the first input number N be arbitrary.)

(f) How many times is the "bypass" from line 5 to line 7 traversed?

(g) How should the program be modified if it is desired to compute the smallest of the N numbers?

9. A very common type of repetitive computation, in which a simple calculation is repeated over and over on different sets of data, is illustrated by Program 6. Assume that each member of a class of N students has taken an experimental measurement of the volume V and the mass M of a sample of a certain substance. Let the input consist of the value N followed by the N pairs of V and M. What does Program 6 then calculate?

PROGRAM 6


START
1. Read input → N
2. K ← 0
3. Read input → V
4. Read input → M
5. D ← M/V
6. K ← K + 1
7. Write K, D
K < N
8. If K < N go to 3
K = N
9.
STOP

10. Now suppose we have a slightly more complicated situation in which each student reports a lower bound V1 and an upper bound V2 for his volume measurement, and a lower bound M1 and an upper bound M2 for the mass. Modify Program 6 so that it will compute for each student the lower and upper bounds for the density implied by that student's data.

What does your program assume about the input data?

11. (a) Write a program which will make a table showing $\underline{n}$, n^2, and n^3 for integral values of $\underline{n}$ from 1 to 100.

    (b) Why does this program apparently have no input?

12. Write a program which will make a table showing $\underline{n}$, $\sum_{k=1}^{n} k$ and $\sum_{k=1}^{n} k^2$ for integral values of $\underline{n}$ from 1 to 20.

13. Construct a program which will, like Program 5, find the largest of N numbers, but which will also produce an integer indicating which of the numbers has the largest value. This integer should equal 1 if the first is the largest, 2 if the second is largest, etc. In case of ties, the integer should indicate the first of the largest values.

14. Construct a program which will find the largest and the second largest of N numbers. Hint: After reading N, read the first number into register A1 and the second into register A2. Then, if $A1 < A2$, interchange the two values, so that it is known that $A1 \geq A2$ (be careful to do the interchange correctly!). Then read the next number into X. If then $X \leq A2$, that value is unimportant and the process may proceed to the next input number. If $X > A2$, then X can replace A2. An interchange of A1 and A2 may now be necessary, as we want A1 always to contain the largest number read to date, and A2 the second largest.

    Use a register named K to count the number of values that have been read in to date. To what value should K be set when it is initialized?

15. A table of loan payments (such as house mortgage payments) is to be prepared. The table is to have five columns: The first is to be the month M (numbered 1 to 12) of each payment, the second the year Y, the third the amount of payment A due at that time, the fourth the interest charge C accrued over the past month, and the fifth the principal P of the loan after that payment.

    The input is to be the month and year of the loan, the total amount of the loan (equal to the principal over the first month), the annual percentage rate R and the amount to be repaid each month. Assume that the same amount is paid each month except in the last month, when a lesser payment equal to the entire remaining principal is made. Construct a program to prepare this table. Assume that each month an interest charge equal to $\frac{R}{12} \times P$ is accrued, and that the monthly payment A is always greater than this. Note that you will have to increment M until it reaches 12, but on the next step M will have to be reset to 1 and Y incremented.

5.3 Basic BASIC

So far we have been discussing how to construct and organize programs. We now consider how to express or "code" a program in a computer language. The candidate languages include BASIC, FORTRAN, ALGOL, PL/I, APL, and perhaps others. We have chosen to use BASIC because it was specifically designed to be used by non-specialists on a time-sharing system,* and as a result is probably the easiest computer language to handle at the start; moreover, computers using BASIC are widely available.

As an example of how to code a program in BASIC we will code Program 4 which computes the mean of N arbitrary numbers. We recall that the input was assumed to consist of the value of N followed by the N numbers to be averaged. Table 5.2 shows this program written both in the symbolic abbreviated English form we have been using, and in BASIC.

You can probably infer most of the rules and conventions of BASIC by examining BASIC programs such as this one, in analogy with learning a natural language by the Berlitz method. However, at the risk of spoiling the fun, we will explain the conventions and rules of the grammar of BASIC.

---

*A time-sharing system is one which has a central computer connected to a number of terminals located at different, convenient places. Each terminal can be used to run programs and a number of terminals can be used simultaneously.

TABLE 5.2
(Program 4 in Symbolic Form and in BASIC)

Symbolic Form	BASIC
1. $S \leftarrow 0$	10 LET S = 0
2. $K \leftarrow 0$	20 LET K = 0
3. Read input $\rightarrow N$	30 READ N
4. Read input $\rightarrow X$	40 READ X
5. $S \leftarrow X + S$	50 LET S = X + S
6. $K \leftarrow K + 1$	60 LET K = K + 1
YES 7. Is $K < N$? NO	70 IF K < N THEN 40
8. $S \leftarrow S/N$	80 LET S = S/N
9. Write S	90 PRINT S
10. Stop	100 STOP
	110 DATA 3, 1.2, 2.6, 3.4
	120 END

BASIC was created with a certain teletype keyboard in mind, and as a result uses only symbols available on that keyboard: letters, numbers, and a few punctuation marks and special signs. No distinction is made between upper and lower case letters. Spaces carry no information and may be inserted for legibility or omitted as one wishes.

In BASIC lines can be numbered with any numbers from 1 to 9999, (from 1 to 99999 on some systems). Notice that in Table 5.2, in the BASIC column, the lines are numbered in increments of 10. The reasons for this curious custom of incrementing line numbers in steps of 10 rather than steps of 1 will be explained when we discuss the secretarial aspects of BASIC time-sharing systems, in which line numbers play an important role.

Now let us discuss each line of Table 5.2. First we have the assignment statement. The first line "$S \leftarrow 0$" translated into BASIC reads "LET S = 0." This means "store the number 0 in register S." Similarly, "$S \leftarrow X + S$" translates into "LET S = X + S," which means "compute the contents of X plus the contents of S and store the result in S." Thus, the apparently self-contradictory statement on line 60, "LET K = K + 1," has the perfectly sensible meaning "add one to the contents of K and store the result back in K," or more briefly "increment the contents of K by one." It is a common complaint that this is a mis-use of the equal sign, but no more suitable sign was available on the teletype keyboard for which BASIC was designed.

Symbols for arithmetic operations may appear to the right of the equal sign in assignment statements, as in lines 50, 60, and 80. Multiplication must always be indicated by an asterisk (*), division by a slash (/), and exponentiation by an arrow pointing upwards (↑), while addition and subtraction are, as you see from Table 5.2, indicated by the usual symbols.

Parentheses may be used as is customary in algebraic expressions. Suppose, for example, that

The contents of A equals 2.
The contents of B equals 3.
The contents of C equals 4.

Then the following BASIC coding lines:

```
220   LET X = B↑2 - 4 * A * C
225   LET Y = 1/A + B
230   LET Z = 1/(A+B)
```

result in the storing of the numbers

$$3^2 - 4(2)(4) = -23 \text{ in X (on line 220)},$$

$$\frac{1}{2} + 3 = 3.5 \text{ in Y (on line 225), and}$$

$$\frac{1}{2+3} = 0.2 \text{ in Z (on line 230).}$$

Output may be handled as on line 90 in Table 5.2, where the instruction "PRINT S" indicates that the contents of register S are to be typed out. One may have several numbers typed out in one PRINT instruction; thus 90 PRINT S, X, K, N would cause the final contents of registers S, X, K, and N to be written out. Note the use of commas to separate the names of registers.

The input numbers are included as part of the BASIC program, on line 110, following the word DATA and separated by commas. These numbers are taken in order, starting at the left, as READ instructions are encountered in the execution of the program.

READ instructions occur on lines 30 and 40. On line 30, in the second column of Table 5.2, "READ N" means "read the next input number and store it in register N." (In this program the number read in by this instruction is the first of the input numbers, following "DATA" on line 110, namely the integer 3.) "READ X" means "read the next input number and store it in register X." If the input data does not fit on one line, several DATA lines are used.

As another example of data input, which incidentally illustrates how several numbers can be read in by a single READ instruction, consider the example

```
1250   READ A, B, C
1260   LET X = B↑2 - 4 * A * C
1270   READ A, F
1300   DATA 2, 3, 4, 8.5, -9.2, 2.1
```

From line 1300 we see that the instruction on line 1250 stores 2 in A, 3 in B, and 4 in C. On line 1270, 8.5 is stored in A (erasing its previous contents, of course), and -9.2 is stored in F. At this point one more number remains ready for input, namely 2.1.

Branching is done with the IF statement. Line 70 "IF K < N THEN 40" means "if the contents of K are less than the contents of N, then transfer control to line 40; otherwise continue as usual to the next line."

Other relations can be used in IF statements. A complete list is given in Table 5.3. Thus "IF W < = Q THEN 850" means "If the contents of W are less than or equal to the contents of Q go to line 850, otherwise continue."

TABLE 5.3

=	equal
>	greater than
<	less than
>=	greater than or equal to
<=	less than or equal to
< >	not equal to

The "END" statement marks the last line of a BASIC program, and "STOP" indicates a point at which computations are terminated.*

One point which, however, needs further comment is the matter of register names. In BASIC these must either be single letters (as A, S, X, Q), a single letter followed by a single numerical digit (as A5, X0, B9, Q4), or a single letter followed by an index enclosed in parenthesis such as A(5) or B(212).

Within these limitations one may name and use a large number of registers in BASIC — thousands, if necessary. However, the first operation involving any register must be to store a number in it, usually by a LET or a READ instruction. Otherwise, one has a "bug" in the program, which involves asking for the contents of a register whose contents have not yet been defined.

This completes our survey of basic BASIC, and covers perhaps one-third of the total vocabulary of BASIC. This is enough to express quite a large class of programs.

Questions

1. A trace of Program 4 was done in Question 2 of Section 5.2. Does this trace apply to the BASIC version of Program 4 shown in Table 5.2?
2. What change is necessary in order to make Program 4, in its BASIC form, average the numbers 4.1, 5.3, 6.7, 9.5?

---

*How to actually run a program in BASIC on a time-sharing terminal will be discussed briefly in the next section.

3. What does each of the following BASIC programs do? (Hint: When in doubt, do a trace.)

(a)

```
10  LET A = 2
20  LET B = 5
30  LET C = 12
40  LET X = B*B-4*A*C
50  PRINT X
60  STOP
70  END
```

(b)

```
10  READ A,B,C
20  LET X = B↑2-4*A*C
30  PRINT X
40  STOP
50  DATA 2,5,12
60  END
```

(c)

```
 20  LET K = 0
 40  READ A,B,C
 50  LET X = B↑2-4*A*C
 60  PRINT X
 70  LET K = K + 1
 80  IF K < 5 THEN 40
100  STOP
110  DATA 2,3,4
111  DATA 5,8,10
112  DATA -2,6,-3
114  DATA 8,9,1
200  END
```

(d)

```
 20  LET K = 0
 40  READ A,B,C
 50  LET X = B↑2-4*A*C
 60  LET K = K + 1
 70  PRINT K,X
 80  IF K < 5 THEN 40
100  STOP
110  DATA 2,3,4,5,8,10
111  DATA -2,6,-3,4,11,-7,8,9,1
999  END
```

(e)

```
100  LET K = 1
110  LET K2 = K*K
120  LET K3 = K*K2
130  PRINT K, K2, K3
140  LET K = K + 1
150  IF K < 101 THEN 110
160  STOP
999  END
```

(f)

```
 10  LET K = 1
 20  LET F = 1
 30  LET K = K + 1
 40  LET F = F*K
 50  PRINT K,F
 60  IF K < = 10 THEN 30
 70  STOP
100  END
```

(g)

```
 40  LET X = 0
 50  LET Y = 1
 70  LET Z = X + Y
 80  PRINT Z
 90  LET X = Y
100  LET Y = Z
110  IF Z < 10000 THEN 70
120  STOP
130  END
```

4. Find any bugs present in the following programs, all of which are supposed to compute 2.5 times (-1.3):

(a)

```
10  LET A = 2.5
20  LET B = -1.3
30  LET C = AB
40  PRINT C
50  STOP
60  END
```

(b)

```
10  LET A = 2.5
15  LET C = A*B
20  LET B = -1.3
25  PRINT C
30  STOP
35  END
```

(c)

```
19  LET A = 2.5
20  LET B = -1.3
25  LET C/B = A
28  PRINT C
30  STOP
95  END
```

(d)

```
51  LET A9 = 2.5
52  LET A10 = -1.3
53  LET CX = A9*A10
54  PRINT CX
55  STOP
56  END
```

(e)

```
 40  LET A = 2.5
 45  LET B = -1.3
 48  LET C = 0
 50  LET C = (A*(B+C)+C)*1
 20  PRINT C
 90  STOP
100  END
```

(f)

```
210  READ A,B
212  LET C = A*B
215  PRINT C
216  STOP
218  DATA 2.5,-1.3,1.3,-4.12,62.5
219  END
```

(g)

```
  40  READ A,B
 250  LET C = A*B
 900  PRINT C
1221  STOP
1222  DATA 2.5
1223  END
```

5. Code the program of Question 6 at the end of Section 5.2 in BASIC. Include input data such that the program will compute the mean and variance of the 10 numbers 1,2,3,4,5,6,7,8,9,10.

6. Code Program 5 in BASIC, using N = 10 input numbers of your choice. (The program is to find the largest of these 10 numbers.) Note: The BASIC for line 9 is 'PRINT "THE LARGEST IS".' (THE LARGEST IS must be enclosed by quotation marks. If it is not, you have a bug because the computer reads this as a four word instruction: PRINT THE LARGEST IS, which is not part of the BASIC vocabulary.)

7. Code Program 6 in BASIC, supplying input data as follows:

Student	V1	V2	M1	M2
1	47	52	112.1	112.6
2	49	54	112.5	112.8
3	46	51	111.9	112.4
4	48	50	112.2	112.5

8. Code the program of Question 13 at the end of Section 5.2 in BASIC, using input data of your choice.
9. Code the program in Question 14 at the end of Section 5.2 in BASIC.
10. Code the program in Question 15 at the end of Section 5.2 in BASIC.
11. Program 5 is a trot giving a program in two different languages. Discuss whether this can be considered analogous to a trot giving the Gettysburg Address in English and in French. What are the points of similarity between the computer language example and the natural language example, and what are the points of difference?
12. Below is Program 4 expressed in FORTRAN, another very much-used computer language. The more cumbersome way in which input and output is handled in FORTRAN, and the fact that FORTRAN distinguishes between two types of numbers, called "integers" and "real numbers," are two of the factors that make FORTRAN somewhat harder to handle than BASIC at first.

    Without trying to understand everything about this FORTRAN program, see, by comparing it with the other versions of Program 4 if you can identify some ways in which FORTRAN is similar to BASIC and some ways in which the two languages differ.

```
          SUM = 0
          K = 0
          READ (5,99) NTOT
     99   FORMAT (I3)
     10   READ (5,98) XNEW
     98   FORMAT (F10.5)
          SUM = SUM + XNEW
          K = K + 1
          IF (K.LT.NTOT) GO TO 10
          SUM = SUM/FLOAT(NTOT)
          WRITE (6,98) SUM
          STOP
          END
 /DATA
 003
 1.2
 2.6
 3.4
```

13. The successive approximation process for computing the square roots (Section 3.6) of some number V generates the following sequence of iterates:

$$x_1 = \frac{1}{2}\left(x_0 + \frac{V}{x_0}\right)$$

$$x_2 = \frac{1}{2}\left(x_1 + \frac{V}{x_1}\right)$$

$$\vdots$$

$$x_{k+1} = \frac{1}{2}\left(x_k + \frac{V}{x_k}\right)$$

Write a program for finding square roots by this method taking $\frac{1 + V}{2}$ as the initial guess.

Use the fact that for each <u>k</u>, the value of $\sqrt{V}$ lies in the interval between x_k and $\frac{V}{x_k}$ to obtain a criterion for terminating the iterations by writing your program so that when $\left|x_k - \frac{V}{x_k}\right|$ becomes less than 10^{-5}, the iterations will be stopped.

5.4 Running BASIC Programs

Let us assume that you have successfully "signed on" at a time-sharing system terminal, with the assistance of a friend or by following instructions posted on the wall, so that you are confronted by a "live" terminal connected to a BASIC time-sharing computer system. Your activity henceforth will consist mainly of typing in "lines" and pressing the "carriage return" key at the end of every line (step). Each time you return the carriage the line of information you have typed in, encoded in some fashion, is ready to go to the computer. Being very fast, the computer is able to look at each terminal several times a second (this is why it is called a "time-sharing" computer), and take from your terminal a message, namely the encoded line of typing, whenever one is ready to be sent.

Eventually you will have typed in your entire program, and the computer will execute it. But until that point is reached, the computer system acts as your personal secretary, taking dictation and frequently putting in its "two cents worth." The computer will have assigned to you a portion of its memory to be your "work space," in which it will record the lines you type in, appropriately encoded. (This will probably be a certain number of "tracks" on a "disk file," which will be identified for you if you take a guided tour of the computing center.)

The computer, of course, is not really doing any thinking on its own — it is slavishly following a very long and elaborate program, which has been written by specialists to control the computer during time-sharing operation. What you type in is, in effect, input data for this "operating system" program, and by means of many branches — IF statements, in effect — the program can test each line you type in, send you an "error message" if a line violates certain conventions of BASIC, or store the line in your work space if it passes all tests.*

---

*We are here describing a system "dedicated" to BASIC. BASIC is also available on some systems not fully dedicated to BASIC; in this case error messages do not occur as you type in each line, but only when you attempt to run your program.

After you have typed in your entire program (or at intermediate points on demand), the computer system sorts the lines by line number — that is, it arranges them in increasing order of line numbers. This is why every line must be numbered. (Try typing in a line without a number, and you will see that the computer rejects it, sending you a message of some sort to this effect.) If the same line number appears more than once, the computer saves only the last line typed in with that number.

This is, in effect, secretarial service, performed for you under control of the operating system program. The implications of this service are as follows:

If you want to correct or change a line, just type in the line you want, with that line number. That will replace the former version.

You don't need to type your program in order from top to bottom. Just use line numbers correctly.

If you want to insert one or more lines between two lines of your program, just type in lines with intermediate numbers. This is the reason for the custom of incrementing line numbers in steps of 10 as in Table 5.2 in Section 5.3; unforeseen insertions are then easy to fit in.

If at any point you want to see what is in your work space, type LIST. This will cause the system to sort the lines in your work spaces and then type them out for you to see.

For example, suppose you type in

```
10   LET A=5
20 LET B = 0
10 LET A = 2
30 LET C = A+B
200STOP
210END
40PRINT C
20 LET B = 3
LIST
```

Then the computer, under control of its operating system program, will clean up the contents of your work space, and type them out; the result will probably look something like this:

```
 10  LET A = 2
 20  LET B = 3
 30  LET C = A+B
 40  PRINT C
200  STOP
210  END
```

Note that the lines have been sorted, the last version of line 20 has replaced the earlier version, and spaces have been inserted according to a conventional pattern.

When you are finally satisfied with your BASIC program, or when you feel like giving it a whirl, type RUN. The computer will thereupon attempt to execute your program. If all goes well, you will see your output appearing on the terminal typewriter; each time a PRINT instruction is encountered in the program, the register contents referred to are sent to your terminal, which types them out. When a STOP instruction is encountered (or when a bug is detected by the system), operation ceases, and you may continue typing input to modify your program. When you are all done you may sign off by typing BYE.

Questions

1. Have someone show you how to use the terminal you will be using. Prepare a sheet for your own future use, which includes notes on how to turn the terminal on, how to sign on, how to save and retrieve programs, how to sign off, and other such useful information.
2. Experiment with typing in a program. Type the lines out of order and observe how the system sorts them whenever you ask for a LIST. Observe how lines may be changed simply by retyping them.

5.5 Debugging a Program

Now, how about bugs? In nature, these come in three families (the phylum is arthropoda, the class insecta, and the order hemiptera). Computer bugs can also be classified into families.

The first type of bug includes those which are recognizable by examination of a single line by itself. These are the least pestiferous as the computer immediately recognizes and rejects such lines, and asks you to try again.* For example, if you type

```
100  LET Y - X + W
```

(accidently hitting the minus sign instead of the equal sign), one of the tests which the system makes on each line of input will fail, causing the operating system program to type out an error message rather than storing the line in your work space.

The second family of bugs includes those which the system does not detect and tell you about until you try to run your program. Here are several examples:

```
(a)   10  LET K = 1                  (b)   10  LET A = 2
      20  LET W = 1                        20  LET B = 3
      40  LET W = W + K                    25  LET C = A + B
      60  PRINT K, W                       40  PRINT, A, B, C
      75  LET K = K + 1                    50  LET E = (A+D)*C
      80  IF K < 10 THEN 30                60  PRINT E
      85  STOP                             70  STOP
     999  END                              99  END

(c)   10  READ A,B                   (d)  100  LET A = 10
      20  LET X = A*B                     110  LET K = 0
      30  PRINT A,B,X                     120  LET A = A - 1
      40  READ C,D                        140  LET B = 1/A
      50  LET Y = X+C*D                   150  PRINT A,B
      60  PRINT Y                         160  LET K = K + 1
     100   STOP                           170  IF K < 12 THEN 120
     110  DATA 5.23,-18.7,2.3             180  STOP
     200  END                            9999  END
```

In these programs every line by itself is a plausible BASIC line, yet bugs are present: In example (a) an IF statement refers to a non-existent line, while line 50 of example (b) refers to a register D whose contents have not yet been defined (because no number has yet been stored in register D). In example (c) insufficient data have been provided, and in example (d) division by zero occurs on the tenth time that line 140 is executed.

---

*As remarked earlier, this service is provided only on systems fully dedicated to BASIC.

In each case the execution of the program will be aborted at the point where the bug first causes trouble. An error message identifying the trouble is sent to the terminal, and the terminal is left in readiness to receive corrections or additions to the program, just as if no RUN had ever been requested.

In example (a) trouble comes immediately; one of the first things the BASIC system program does when you type RUN is to check the transfers of control, so that in example (a) no calculations will be made. In the other three examples, however, some calculations will take place, and some output of the program will be obtained before the bug causes the computations to be aborted.

Precisely what will happen in each case depends on the system. Sometimes rather than terminating calculations, a warning message is typed out, but the calculations are allowed to proceed. In this case you will have to deduce what the system did to get around the difficulty if you are to make use of the results of the computation.

The error message one receives usually makes it easy to spot and correct bugs. Sometimes, however, it can be quite difficult to locate them, and detective work is required. In that case temporary insertion of extra PRINT instructions, to yield a partial trace of the calculations, is often helpful in localizing the trouble. In this way the computer can be used to help in the debugging.

The third family of bugs are those that produce programs which run, but just don't compute what you want to compute. As a trivial example, suppose you want to compute $\frac{3.52}{7.71 - 1.98}$ and to that end write the program

```
10   READ A,B,C
20   LET X = A/B-C
30   PRINT X
40   STOP
50   DATA 3.52,7.71,1.98
60   END
```

The computer will compute $\frac{3.52}{7.71} - 1.98$ rather than the result desired. In this case the bug, which was failure to use parentheses in line 20, cannot be detected in the compiling and running of the program.

Perhaps the best protection against such bugs is to run test cases; that is, to run your program with test values of the input numbers, and compare the numbers the computer generates with independently computed answers.

Finally, to close this brief glimpse into life with the computer, we should mention the other secretarial services provided by BASIC. These vary from system to system. However, there should be some means of storing programs in a users' library of the system, usually by typing SAVE. In order to do this you have to give your program a name — on some systems you will have that done when you start typing it in. Then later you can retrieve your program from the library and continue working with it. This is obviously a big help if the program is long, or you are a slow typist. Also, it allows you to use other people's programs. Finally, by typing in SCRATCH, or PURGE, or KILL followed by the name (find out which applies for your particular system), you may remove the program from the users' library; it is important to do this as otherwise the library becomes glutted with old programs.

Some further important secretarial services come under the heading of EDIT operations. For example, it is possible to extract portions of a program in the library, or to delete portions. It is possible to combine together a number of programs or portions of programs into one long program. It is possible to resequence line numbers. At first you won't need these editing services of the system, but later when you start writing long programs and combining subprograms together, they will come in very handy.

Questions

1. Determine by experiment what your particular system does when you type in various incorrect lines of BASIC (line number missing, inadmissible register names, misspellings of words such as LET or PRINT, etc.)
2. Four examples of programs with Type "2" bugs were given in the text. Determine by experiment what happens when you try to run these programs on your system. Would the behavior of the system enable you to locate the bug in each case?

3. Type in and run selected BASIC programs you have prepared to date.

4. (a) Type in and run the square root program you coded in Question 13 of Section 5.3.

   (b) Test this program with several input values V. For some values of V that you use, what is the approximate per cent error in the square root which the program computes?

   (c) Modify your program so that it types out every iterate X_1, X_2, X_3, which it generates. Observe this sequence of iterates for several test cases, and comment on the manner in which the sequence converges to the answer.

Chapter 6. GRAPHS

6.1 Functions; Independent and Dependent Variables

Some of the most powerful applications of mathematics are those dealing with change and with relationships between changing quantities. As an example consider Table 6.1 which is a record of a temperature sounding taken at Washington, D.C. during the early morning hours of August 15, 1936.

TABLE 6.1

Elevation (ft)	Temperature (^oF)	Elevation (ft)	Temperature (^oF)
20	79	5000	67
1000	74	6000	65
2000	76	7000	59
3000	73	8000	56
4000	70	9000	52

The data are given in two corresponding columns. The numbers in the right-hand column refer to the atmospheric temperatures while those in the left-hand column refer to the corresponding elevations. Neither of the two columns taken by itself is at all useful. However, taken together they convey information about the relationship between changes in elevation and corresponding changes in temperature. A table such as the above is one way a relation may be represented.

We may also state relations in words by describing the conditions we impose upon the quantities involved. Consider the following: "For each throw of a die record the value on the side facing up." This is a perfectly good relation between the value on the face of a die and the ordinal number of the throw.

Perhaps the most common way of describing a relation between two quantities is to write an equation connecting these quantities. For example, $A = \pi r^2$ expresses a relation between the area $\underline{A}$ and the radius $\underline{r}$ of a circle.

In words, it states that the area of a circle is equal to its radius squared times the constant π.

In the relation $A = \pi r^2$ the symbols A and r are called variables since they are used to represent many numeric values. In practice when we are dealing with a relation such as the above we usually choose a value for r and then compute the corresponding value of A. That is, we usually think of A as being determined by r, or dependent on r. Therefore we call A the dependent variable and r the independent variable. More generally, the dependent variable is the variable whose values are obtained after values of the independent variable are chosen. These values of the dependent variable may be computed as in the case of the area of a circle or they may be the results of measurements as in the case of the data in Table 6.1. There, elevation, the independent variable, was varied experimentally and atmospheric temperature was measured for the corresponding elevations.

In many situations as described above, it is convenient to think of one variable depending on the other rather than the reverse. For example, it is more natural to think of temperature as depending on elevation than of elevation depending on the temperature. Consequently elevation is chosen as the independent variable with atmospheric temperature becoming the dependent variable. In other cases the relationship between two variables is symmetric in nature and we may arbitrarily choose the independent variable. For example, it is just as natural to say that the area of a square depends on its perimeter as to say that the perimeter depends on the area.

When a relation between two variables is such that for each value of the independent variable there is only one value for the dependent variable, the relation is called a function or sometimes a functional relation. All the permissible values of the independent variable comprise the domain of the function whereas all the values of the dependent variable comprise the range of the function. Thus, for example, in the functional relation in which the value on the side facing up on a die is a function of the ordinal number of the throw, the domain consists of all positive integers while the range is restricted to the integers from 1 to 6.

Giving the values of the independent and dependent variables in numerical form is not the only way of describing a function. The values of the two variables can also be described in graphical form using coordinates in a rectangular coordinate system. Figure 6.1 is such a graphical representation of a typical electrocardigram. It provides a comprehensive view of the variations in voltage as a function of time, much more revealing than could be obtained from any tabulation of corresponding values. For this reason we shall discuss graphic presentations of functions intensively in this chapter.

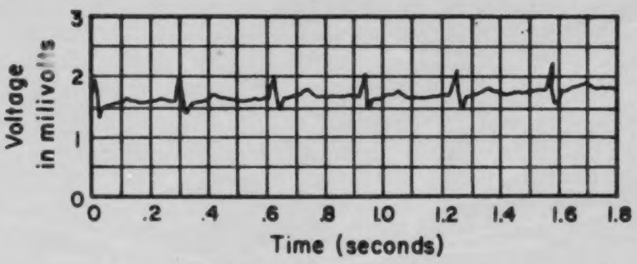


Fig. 6.1

Questions

1. If each of the following statements expresses a functional relation between the variables indicated, which of the variables would most logically be chosen to be the independent variable?

   (a) The day of the month and the corresponding maximum outdoor temperature.

   (b) The atmospheric temperature and the position of the sun in the sky on a sunny day.

   (c) The volumes of spheres and their corresponding circumferences.

   (d) The volumes of spheres and their corresponding surface areas.

2. A useful categorization of variables is in terms of the values which they can assume. Sometimes the variables take on discrete values each separated by some finite difference. Often, however, they take on all the values contained in an interval on the number line.

   (a) Can you give an example of a function whose independent variable takes on discrete values and whose dependent variable takes on all values in an interval.

   (b) Give an example of a function whose domain consists of all values in an interval and whose range has discrete values.

6.2 Choosing Scales for Axes

When a function is graphed we usually plot the independent variable horizontally and the dependent variable vertically. Thus in Fig. 6.1 time appears as x-coordinates or abscissas and is the independent variable while voltage appears as y-coordinates or ordinates and is the dependent variable.

If we are graphing data from a table, the first step is to choose the size of the scales, that is, how large an interval will be represented by each pair of horizontal lines of the graph paper and by each pair of vertical lines. Figure 6.2 represents a graph of the data of Table 6.2. Each division on the vertical axis represents a five-year interval. Obviously, this is not the only possible choice. The same data are plotted in Fig. 6.3 using different scales; here one division on the vertical axis still represents five million people while one horizontal division represents a 10-year interval. Neither graph is incorrect, but the one in Fig. 6.2 has advantages over the other. If we use a scale like the one shown in Fig. 6.3, on a whole sheet of graph paper,the graph will huddle on a small part of the page, leaving most of the area blank and therefore devoid of information. A more expanded scale like that in Fig. 6.2 makes it easier to plot and read the graph accurately.

TABLE 6.2

Population of the United States, 1790 - 1950
From the Statistical Abstract of the United States

Year	Population (millions)	Year	Population (millions)
1790	3.929	1900	76.094
1800	5.308	1905	83.820
1810	7.240	1910	92.407
1820	9.638	1915	100.549
1830	12.866	1920	106.466
1840	17.069	1925	115.832
1850	23.192	1930	123.077
1860	31.443	1935	127.250
1870	39.818	1940	132.594
1880	50.156	1945	140.463
1890	62.948	1950	152.271

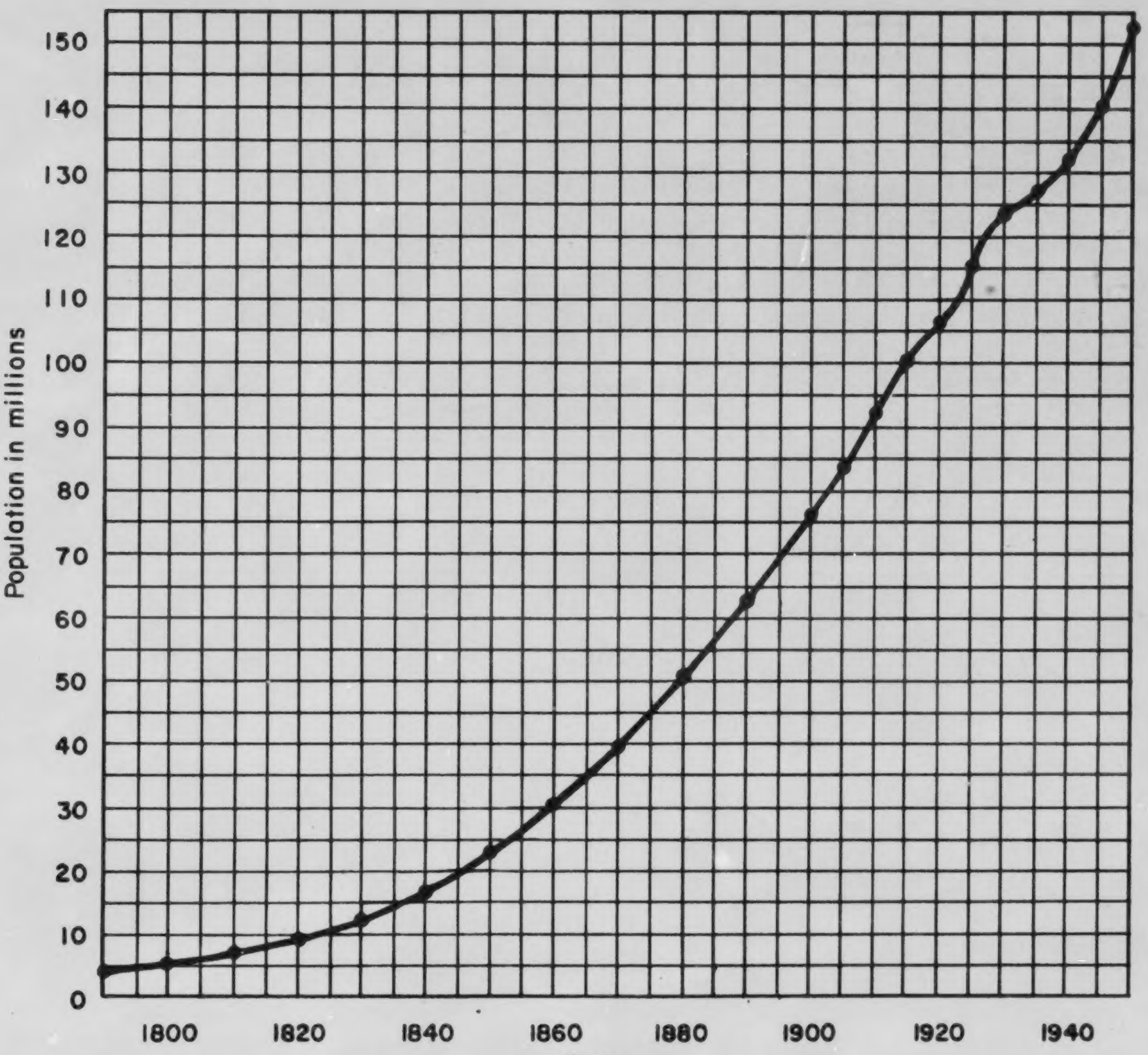


Fig. 6.2. Population of the United States, 1790 - 1950

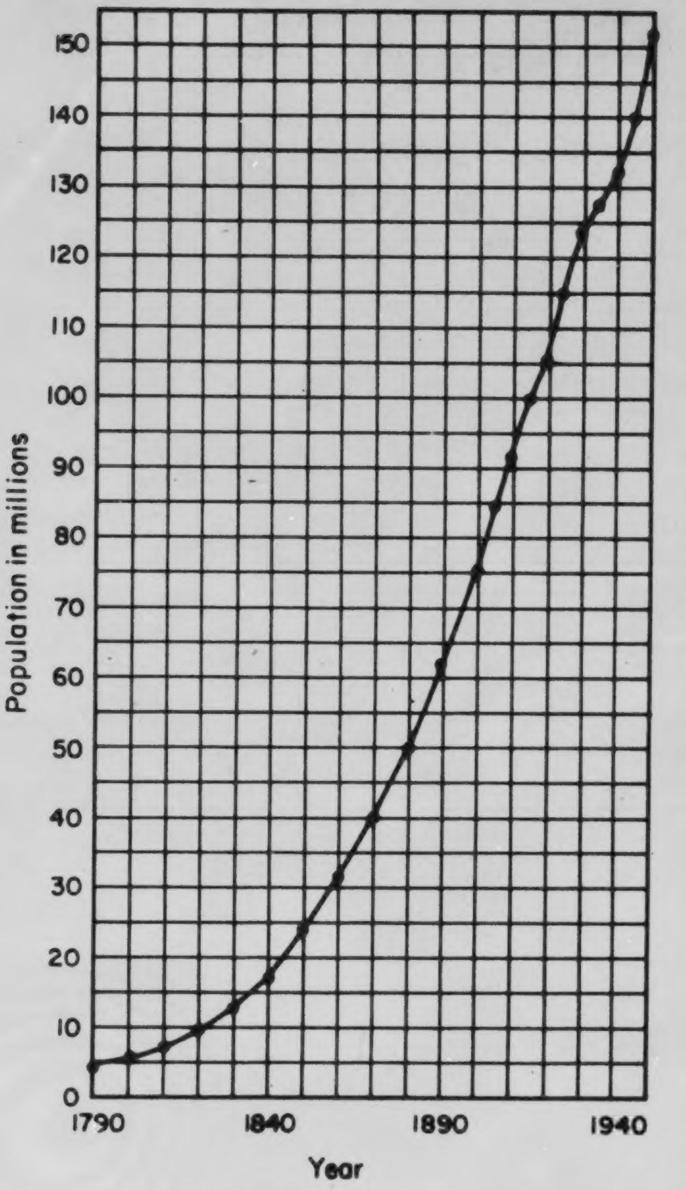


Fig. 6.3. This graph presents the same data as in Fig. 6.2, but plotted using a different horizontal scale.

The population data run from 4.0 millions to 152.5 millions,* and the rulings from zero to 160, well below the top of the page. The graph could be made to cover most of the whole page by using 40 divisions of the same size as those shown in the figure and letting each represent $\frac{152.5 - 4.0}{40} = 3.7$ millions instead of 5 millions. Figure 6.4 shows a portion of the graph of Fig. 6.2 but with each of the original vertical scale divisions representing 3.7 millions and starting from 4.0 millions. Such a vertical

---

*Since we cannot plot points on the graph to better than about the nearest half million, we have rounded off the population data from which we constructed the graph.

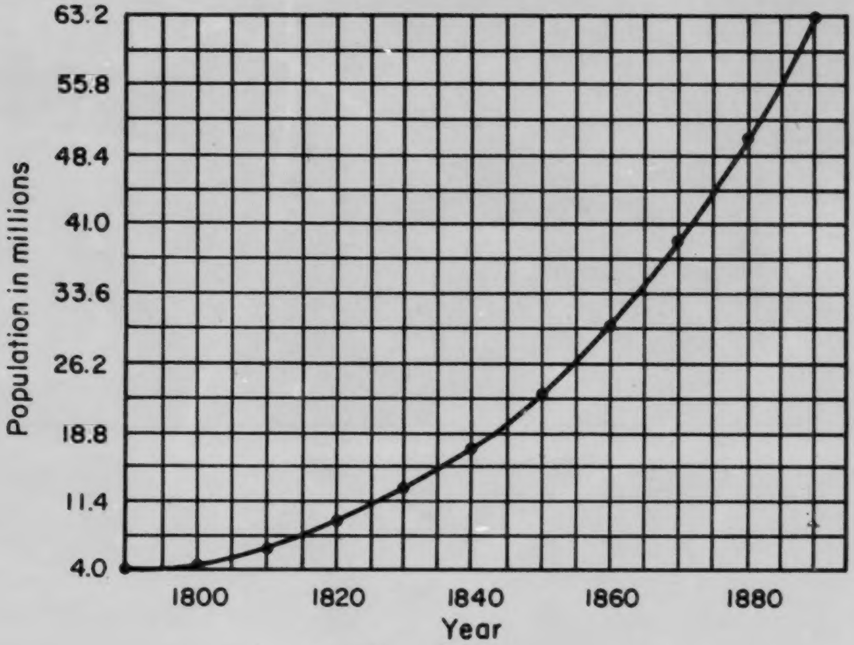


Fig. 6.4

is perfectly legitimate, but it makes plotting and reading the graph laborious. When the scale runs from zero to 160 millions, as in Fig. 6.2, in intervals of 5 millions, the date at which the population was 25 millions, for example, is found easily. Since a graph is intended to be a clear visual display of data, an effort should be made to make it easy to read. Generally, one should choose the interval represented by one division so that the graph has simple decimal scales on which decimal fractions can be plotted and read easily (the scales on a commercial slide rule are examples of this).

If zero on one or both of the scales is not included on a graph, the graph may be misleading if one does not keep in mind where a "missing" zero is (somewhere off the paper). For example, the pressure changes in Fig. 6.5(a) appear to be very large. Figure 6.5(b), however, which includes zero pressure shows that these changes are, in fact, small. (The difference in the two graphs is analagous to comparing numbers by their absolute difference and by their percentage difference.)

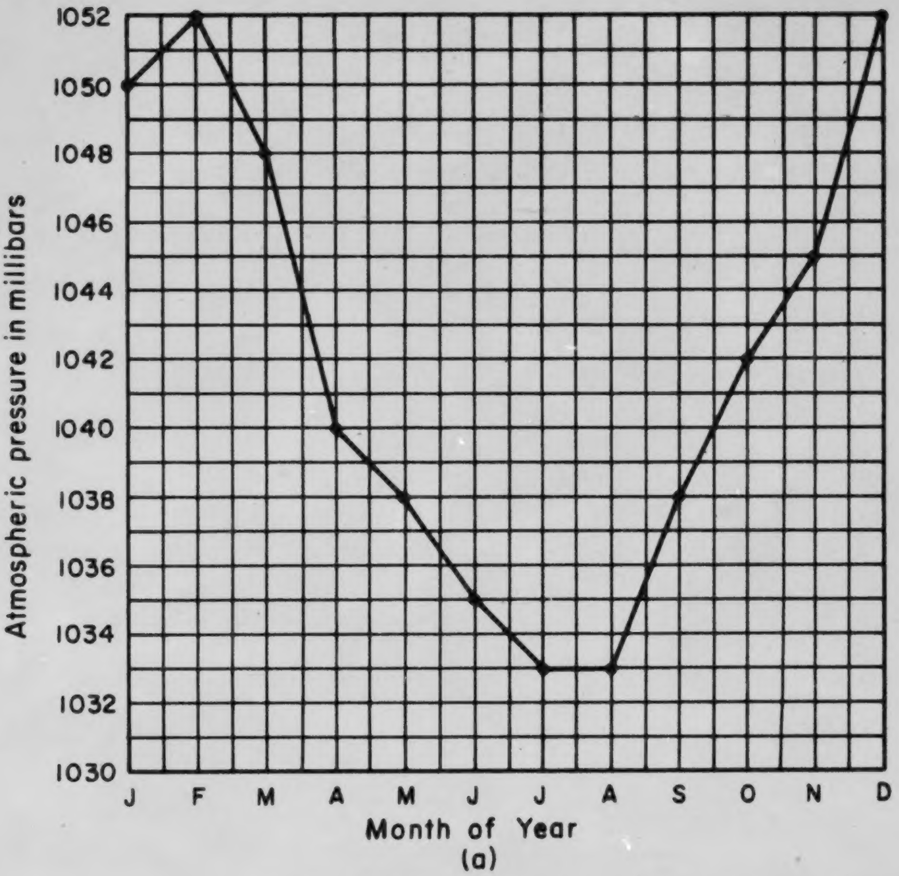
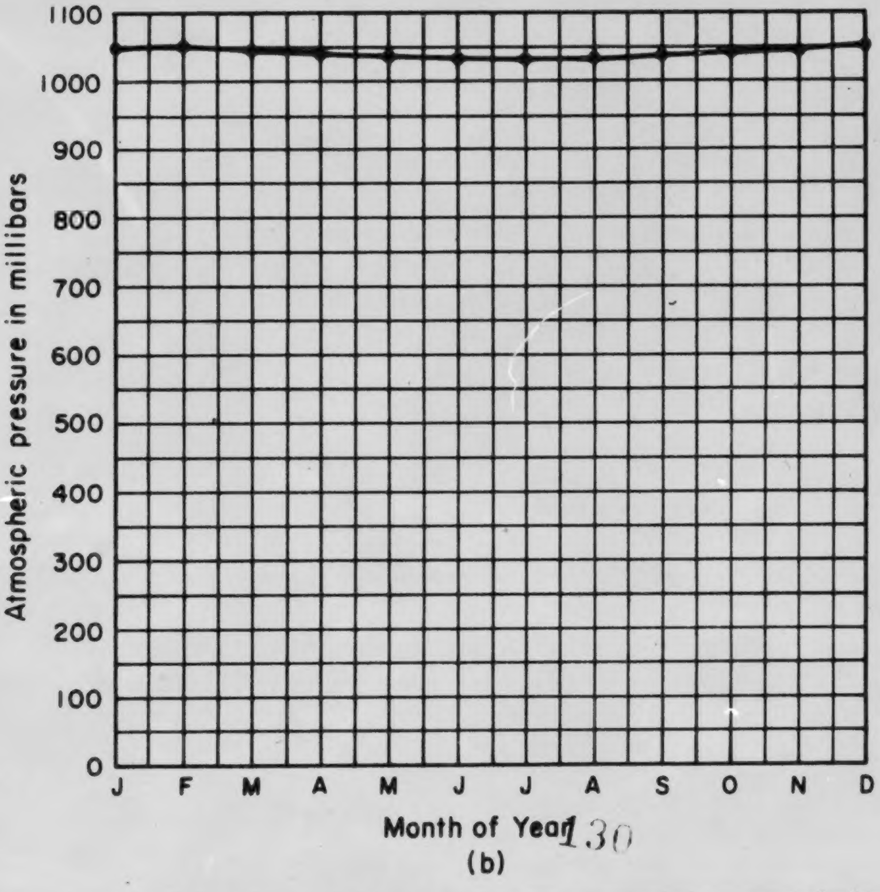


Fig. 6.5. Highest recorded atmospheric pressure each month in New York City during the period 1869 - 1963.



Whether one includes zero or not depends on the purpose for which the graph is drawn; there is no general rule. A graph like that in Fig. 6.5(a), for example, can be misleading to someone who sees such a graph for the first time. (A climatologist, who is often concerned with small pressure changes, uses such graphs all the time and is not misled.)

Sometimes there is no question about what should be done. Suppose you are taking temperature readings once every minute of a substance as it cools to room temperature. You can start your graph at time equal to zero or at the actual time your watch shows when you start taking readings. But it would be pointless to start at temperature equal to zero, since you know the temperature will not fall below room temperature. In this case, room temperature is the best choice for the origin of the ordinate scale.

Questions

1. Figure 6.6 contains two graphs on one piece of graph paper. The lower curve is a plot of the time of day that Venus rose throughout 1968, and the upper curve shows the times Venus set, in the same year.
   (a) On what date did Venus rise earliest?
   (b) On what date did it set latest?
   (c) On what date was it above the horizon longest?
2. (a) Use Fig. 6.4 to find the population in the years 1810 and 1840.
   (b) Repeat (a) using Fig. 6.2.
   (c) Are the points easier to locate in (a) or (b)?

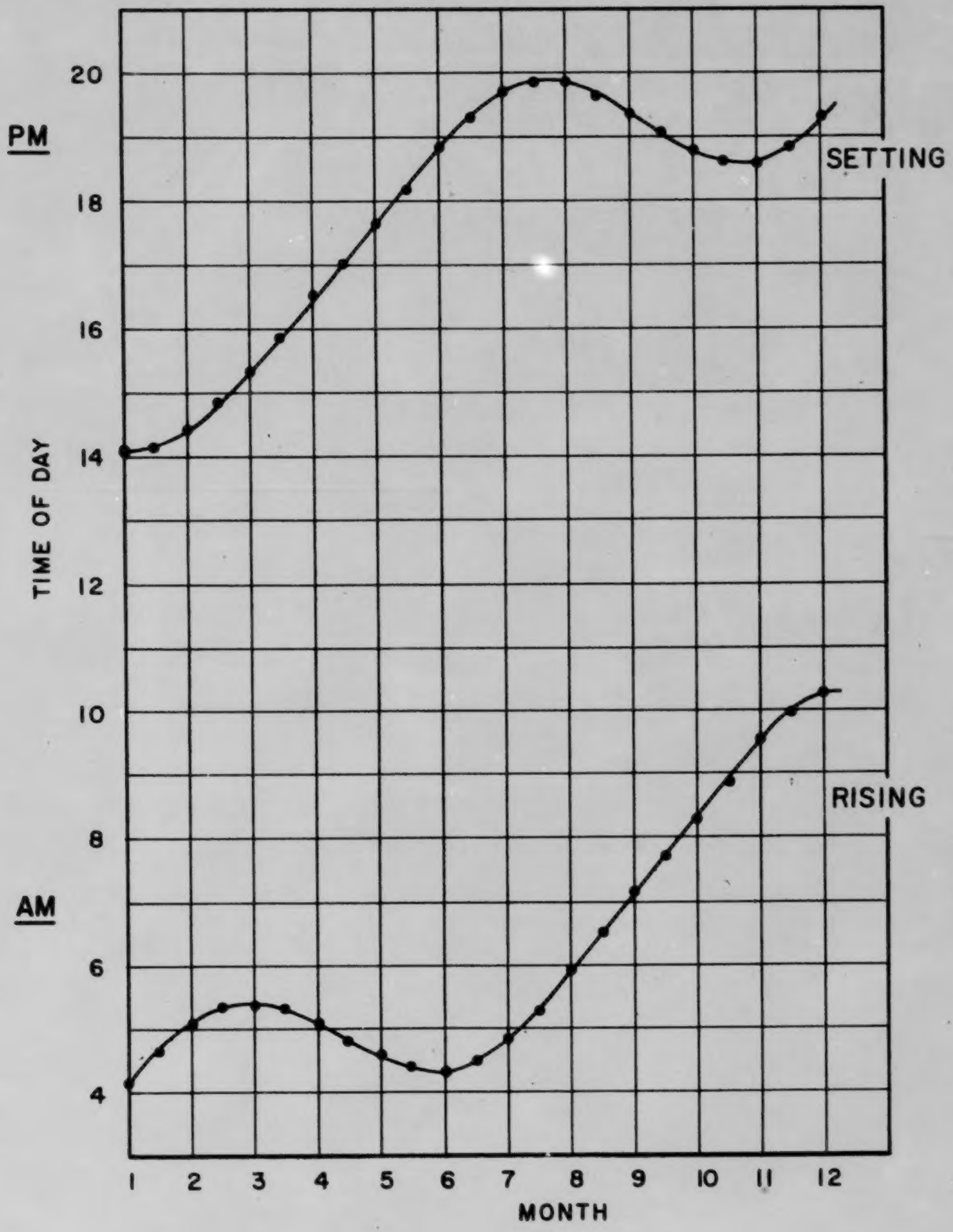


Fig. 6.6. Rising and setting times for Venus in 1968. ("1" on the horizontal scale represents Jan. 1; "2" represents Feb. 1, etc.)

3. Figure 6.7 shows three possible scales for a graph. On each one, locate the points: 0.25, 1.7, 1.8, 2.5, 0.33. Are all three scales equally easy to use? If so, why?

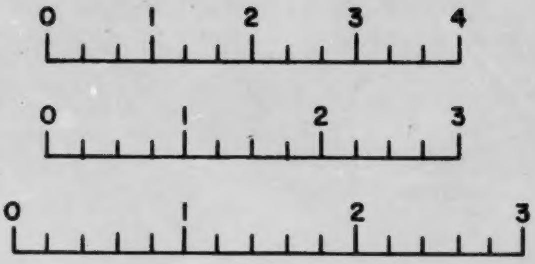


Fig. 6.7

4. Label or describe scales suitable for graphing the following sets of data. Make sure, not only that all the data described can fit on the graph, but also that interpolation is made easy — that the smallest divisions correspond to reasonable numbers.

   (a) Height between 2 and 6 feet
   Age between 0 and 17 years

   (b) Public debt between 240 and 380 billion dollars
   Years between 1950 and 1966.

   (c) Fahrenheit temperature between 32° and 212°
   Centigrade temperature between 0° and 100°

   (d) Day of year between 0 and 365
   Time of sunrise between 4:13 and 7:39

5. The table below gives the masses of steel spheres of different diameters. Draw a graph of the data.

Diameter (cm)	Mass (gm)	Diameter (cm)	Mass (gm)
0.20	0.03	1.20	7.42
0.40	0.27	1.40	11.76
0.60	0.93	1.60	18.00
0.80	2.20	1.80	25.00
1.00	4.30		

6. The table below gives the masses of spheres (made of a more dense material than iron) for different diameters. Plot these data on a graph.

Diameter (cm)	Mass (gm)	Diameter (cm)	Mass (gm)
0.20	0.06	1.20	14.84
0.40	0.54	1.40	23.52
0.60	1.86	1.60	36.00
0.80	4.40	1.80	50.00
1.00	8.60		

7. Compare your estimates of (i) the absolute uncertainty and (ii) the relative uncertainty in determining the change in maximum average pressure from February to June in both Fig. 6.5(a) and 6.5(b).

6.3 Smooth Curves and Uncertainty

Figure 6.8 presents the data of Table 6.1 in graphic form. The lines drawn between data points enable us to estimate the temperature at altitudes

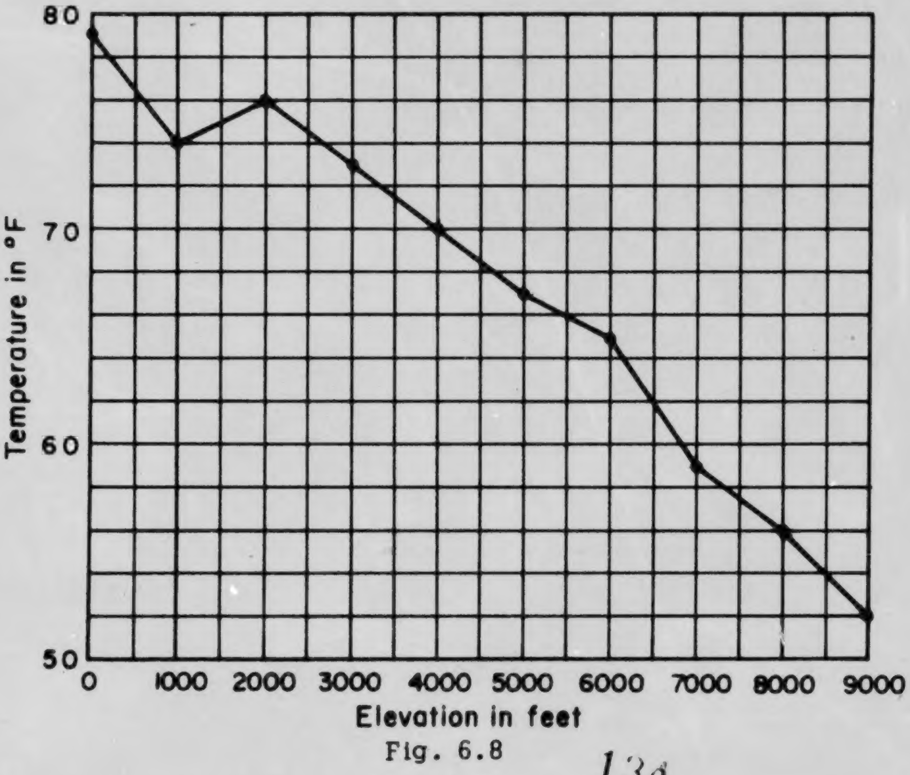


Fig. 6.8

other than the ones at which measurements were taken. In choosing what sort of line to draw on a graph of known data points, one has a wide choice. Using straight lines, as in Fig. 6.8, is a simple choice but not necessarily the most reasonable. Note that the lines joining successive data points meet at angles, forming corners all along the graph. If the measurements of temperature had been made at altitudes other than those listed in Table 6.1, the data points would appear at other places on the graph than on the lines, and consequently lines between these points would meet at corners in places other than those of Fig. 6.8. The corners have no significance in the physical relationship of the temperature to elevation, since the temperature changes in a smooth, regular fashion best described by a graph that is a smooth curve.

By drawing a smooth curve that includes the points in Fig. 6.8, we can connect them so that there are no corners. This may be done freehand or with the aid of a French curve (a plastic template with many different curves which may be fitted against the points on the graph to make a smooth curve). A smooth curve, like that drawn in Fig. 6.9, is not unique, but de-

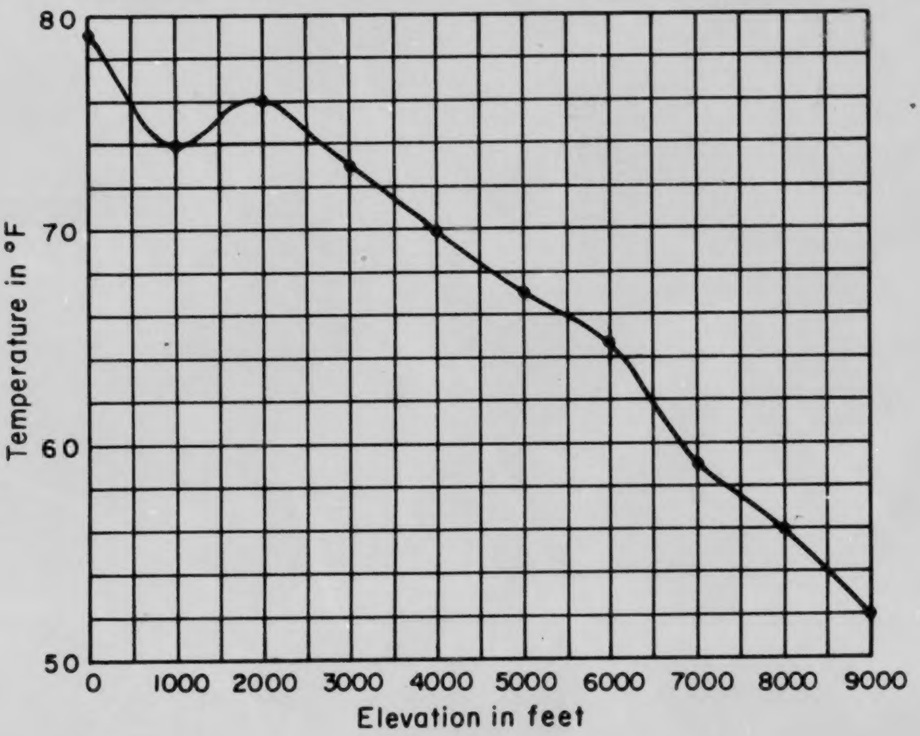


Fig. 6.9

pends on the judgment of the person drawing the curve. The curve may or may not pass through the points that would result from additional measurements, but it is more likely to do so than a series of straight lines connecting the points as in Fig. 6.8. The question raised whenever a line is drawn through a finite number of data points on a graph is how closely it approximates the physical situation being represented. The greater the number of data points in a given interval, the more accurate the graph is likely to be. That is, if the temperature had been measured at intervals of a foot instead of 1000 feet, the points plotted on the same scale as Fig. 6.8 (or Fig. 6.9) would run together and appear to form a continuous smooth curve on the graph, closely approximating the actual physical situation.

So far, in discussing smooth curves we have assumed that the uncertainty in the data is smaller than the uncertainty in actually plotting the data. If the uncertainty in the measurements for data points is larger than this, we must take it into account in plotting a graph. In Chapter 1 we represented an uncertainty in a physical number by an interval on the number line. If we replace a point on each axis by an interval, we replace a point in the plane by a rectangle.

Figure 6.10 is a graph drawn without taking uncertainties into account. It was made from a table of data for the mass and the corresponding volume of a metal. We have drawn a smooth curve through all the points just as we did in Fig. 6.9. However, if we take into account the uncertainties in the measurements (the mass was measured very roughly with an uncertainty of ±5 gm and the uncertainty in the volume was ±0.5 cm^3), the data are consistent with a straight line, as shown in Fig. 6.11. Note that the straight line passes within the uncertainty rectangles whose sides are 10 gm and 1.0 cm^3. Of course, the wiggly curve in Fig. 6.10 is also consistent with the data. But whenever possible we try to fit data with the simplest possible curve. (Occasionally, however, more refined measurements show that an earlier and simpler curve is only an approximation of the relation between the quantities.)

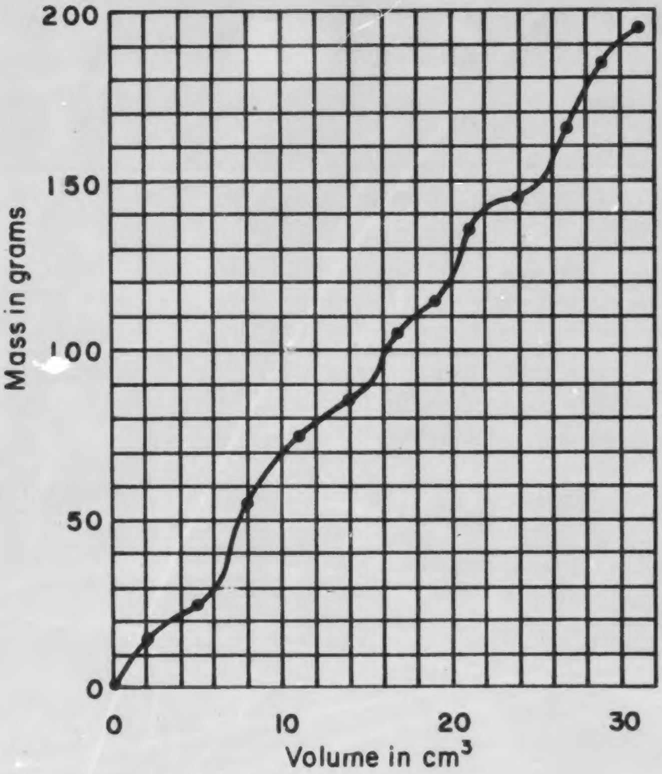
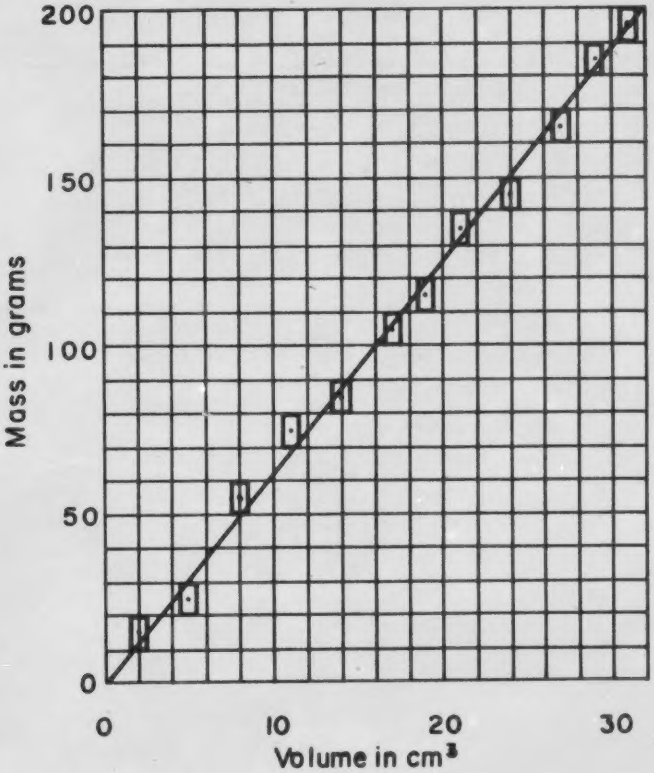


Fig. 6.10

Fig. 6.11



We do not always draw the uncertainty rectangles on a graph, but in deciding how to draw a curve through a set of data points the approximate size of the uncertainty rectangles must be kept in mind.

Questions

1. The following problem is best solved using an electronic desk calculator or computer:

For those who watch the stock market, the Dow-Jones Industrial Average is vital information. From issues of the Wall Street Journal, here are a few days' quotations:

Dow-Jones Industrial Average (November 1969)

	Open	11:00	12:00	1:00	2:00	Close
*Nov.7	856.19	859.75	860.22	860.61	860.94	860.48
Nov. 10	862.00	863.45	865.69	865.48	863.52	863.05
Nov. 11	861.07	861.01	858.96	858.83	859.23	859.75
Nov. 12	858.57	857.91	858.43	857.97	857.91	855.99
Nov. 13	853.15	852.69	850.51	849.52	850.45	849.85
*Nov. 14	849.19	864.88	846.55	847.45	849.06	849.26
Nov. 17	846.36	844.24	843.65	843.26	842.99	842.53
Nov. 18	840.81	841.21	842.20	842.79	843.19	845.17
Nov. 19	845.53	843.26	841.80	841.00	840.62	839.96

---

*Friday

(a) Take an average value of the Industrial Average for each day and plot it with the date. (Remember to include week-ends when marking divisions on the axis.) From the spread of the numbers for each day, estimate an uncertainty and use uncertainty lines on the graph. Draw a smooth curve through the lines.

(b) If November 15 had been a trading day, what might have been the Industrial Average?

(c) Can you make a similar guess about the possible average for November 9? Why, or why not?

2. The table below gives the masses of different volumes of an alloy. The uncertainties were: mass, ±5 gm; volume, ±0.5 cm^3. Draw a graph of the data including uncertainty rectangles.

Volume (cm^3)	Mass (gm)	Volume (cm^3)	Mass (gm)
1	15	10	95
3	25	12	115
5	45	13	125
7	65	15	155
8	85		

3. The table below gives the volumes of spheres of different diameter. Draw a graph of the data.

Diameter (cm)	Volume (cm^3)	Diameter (cm)	Volume (cm^3)
0.6 ±0.1	0.2 ±0.1	2.1 ±0.3	5.0 ±1
0.8 ±0.1	0.35 ±0.1	2.5 ±0.3	7.0 ±1
1.0 ±0.3	0.45 ±0.1	2.7 ±0.3	8.0 ±1
1.1 ±0.3	0.8 ±0.1	3.0 ±0.3	12.5 ±1
1.4 ±0.3	1.0 ±1		

4. During an experiment with gases, air was allowed to flow past a heater in a tube, and the temperature of the air leaving the tube was measured at various times. The data are tabulated below.

TABLE 6.3

Temperature (°C)	Time (sec)
23.6	30
24.7	85
27.3	210
28.3	305
29.4	370
30.0	430
30.6	490

The uncertainty in the temperature measurements is ±0.1 degree. The time is measured to within ±5 seconds. Plot the data with uncertainty rectangles and draw a reasonably smooth curve through them. How distorted would the curve have been if you had tried to draw a curve that exactly passed through all the points?

6.4 Interpolation and Extrapolation

You already have some experience interpolating on graphs — determining the values of variables between data points or between division marks. (We did this in Section 4.5 to make a convenient decimal scale for a power-of-ten slide rule and you have interpolated between divisions on the graphs in this chapter.)

Sometimes linear interpolation (interpolating on a graph that has straight lines connecting the data points) is as good as interpolation from a smooth curve, but not usually. Table 6.4 gives the distances that can be seen over the ocean from various heights above the water. These data are plotted in Fig. 6.12.

TABLE 6.4

Height (feet)	Distance (miles)
0	0
10	3.9
50	8.7
100	12.3
150	15.1
200	17.4

Suppose you want to know the distances visible from heights of five feet and 120 feet. First, from the smooth curve you can read values of about 2.7 miles and 13.5 miles respectively. To compare these numbers with interpolation from a line graph, we can use straight lines between the points for zero and 10 feet, and between the points for 100 and 150 feet (dashed lines on the graph). Using the lines for interpolation, one gets 2.0 miles and

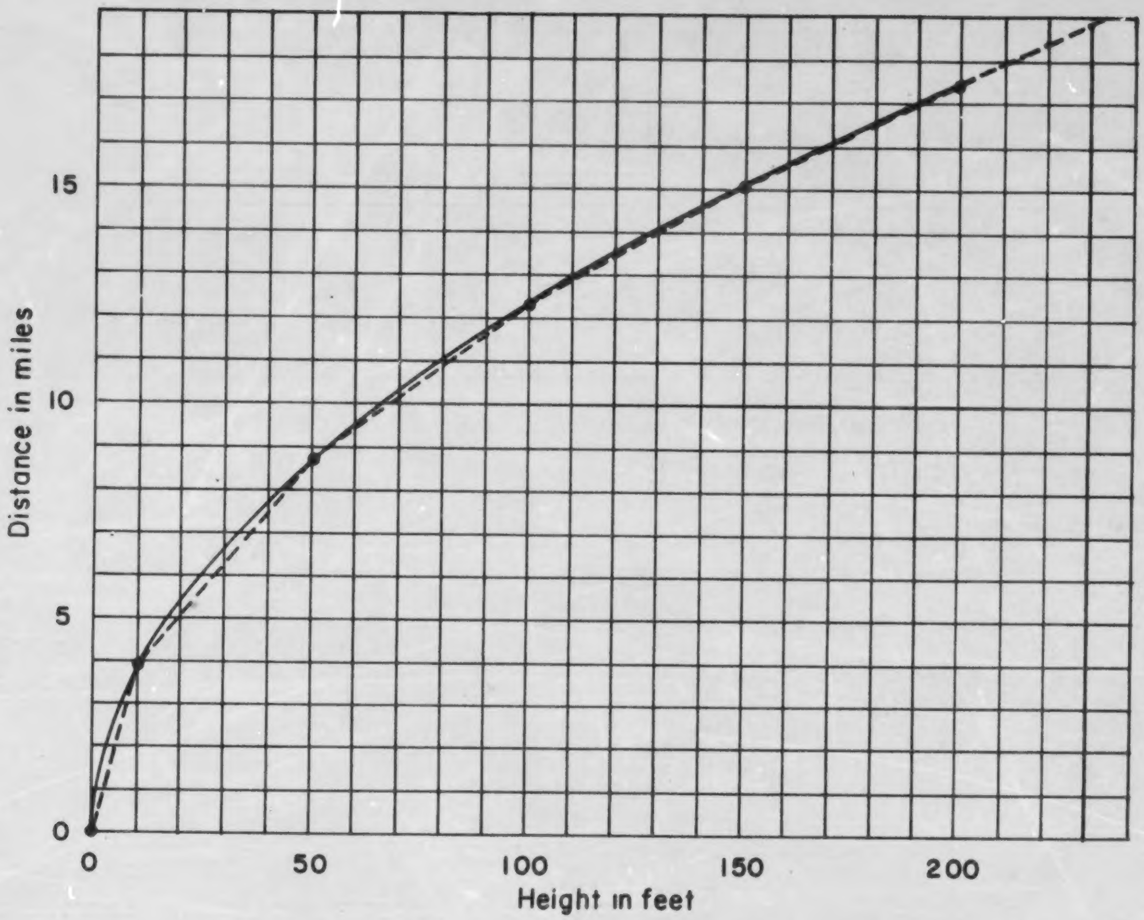


Fig. 6.12

13.4 miles. This linear interpolation is 26 per cent low at five feet, and 1 per cent low at 120 feet compared to interpolation on the smooth curve.

The drawing of a smooth curve through or close to many data points allows us to take account of several adjacent points at once in deciding how curved to make the segments between points, while a straight-line segment is determined by two points only. Thus, interpolation by a smooth curve uses more than just two pieces of information.

It is worth noting that interpolation in decimal fractions is much easier if it is done on a graph with a decimally divided scale, as you found out when you interpolated on the graph in Figs. 6.2 and 6.4 in answering Question 2 of Section 6.2.

The idea of interpolation can be extended to estimating values of variables outside the limits of the known points by extending the curve a short distance beyond those limits, and these can then be used to make estimates. This process is called extrapolation. In Fig. 6.12, for example, the dashed line extending past 200 feet is an extrapolation of the curve. The further one ventures from the known data, the more the curve deviates from the straight line, and the errors in extrapolation increase.

Both interpolation and extrapolation should be applied with caution. Extrapolation involves venturing into unknown territory beyond known points and should not be trusted far from the known data. Interpolation, finding values between known points, seems to be safer. Not all variables inspire this confidence, however.

In Fig. 6.13 the size of the U.S. Army plotted at 10-year intervals gives the solid curve. The size seems to increase smoothly with time. If, however, intermediate points are plotted (x's), the dashed curve results and the enormous effects of World War II and the Korean War become evident. In this case, 10-year intervals are too large to provide an accurate graph.

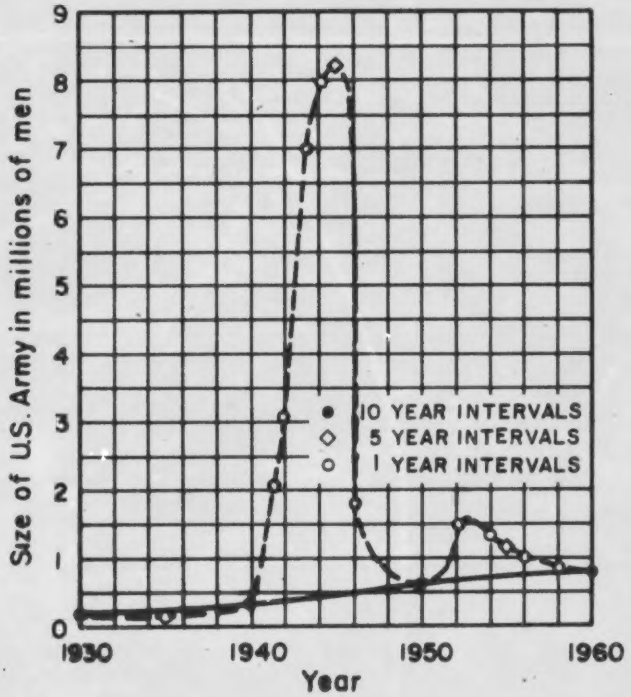


Fig. 6.13

Questions

1. In the third paragraph of Section 6.4 it is stated that the "linear interpolation is 26 per cent low at five feet and 1 per cent low at 120 feet." In terms of Fig. 6.12, how are the figures 26 per cent low and 1 per cent low arrived at?

2. From a few calculations of the volume $V = \frac{4}{3} R^3$ of a sphere (where R is the radius) you can plot the volumes and the corresponding radii and then use the graph to read the volume directly for any value of R. Table 6.5 gives a few values for V and R.

TABLE 6.5

R (cm)	V (cm^3)	R (cm)	V (cm^3)
0	0	1.25	7.24
0.25	0.07	1.40	11.49
0.50	0.52	1.60	17.16
0.75	1.77	2.00	33.52
1.00	4.19		

(a) Plot the points and draw a smooth curve. From this graph, read off values of the volume for radii of 1.10 cm, 1.50 cm, and 1.80 cm.

(b) In which regions of the graph would linear interpolation be reasonable?

3. Using the data points in Table 6.2 draw a graph of the population of the United States during the years 1920 through 1950 and extrapolate it (using a French curve) to estimate the population in 1980 and 2000. How do your estimates compare with those of your classmates?

4. An experiment is done in which a container of water is heated, and the temperature read every 0.1 minute. A table of the data is:

Time (minutes)	Temperature (°C)	Time (minutes)	Temperature (°C)
0.0	29.4	0.6	37.2
0.1	30.7	0.7	38.3
0.2	31.9	0.8	40.0
0.3	33.2	0.9	40.8
0.4	34.5	1.0	42.1
0.5	35.8	1.1	43.4

(a) Make a graph of the data. Considering the accuracy to which the measurements are given in the table, estimate the size of the error rectangles. Are they large enough to be significant on the scale of your graph?

(b) Connect the points with a smooth curve. Does one point appear to be out of line? Draw a better curve through all the points but that one. If that point is actually in error, how much is the smooth curve including it distorted in comparison with the curve not including it? If the apparently "wrong" point is discarded, what is a reasonable guess for the temperature of the water at that time?

(c) What would you expect the temperature to be at the end of 1.2 minutes? At the end of 2.9 minutes?

5. Most curves, viewed under sufficiently high magnification, appear to be straight-line segments over the field of the magnifier. A similar magnifying effect can be obtained by plotting the part of the curve that was magnified on a graph where the divisions of the graph paper represent very small increments of the variables. This can be demonstrated quite simply by plotting the squares of numbers for several choices of scale:

(a) For numbers from 0 to 2 plot the squares of the numbers from 0 to 2 on the vertical axis, choosing x-coordinates (the independent variable) on the horizontal axis at intervals of 0.20.

(b) Plot the squares of numbers from 0.4 to 0.6 with x-coordinates at intervals of 0.02.

(c) Plot the squares of numbers from 0.48 to 0.52 with x-coordinates at intervals of 0.005.

(d) Using a straightedge as a standard of comparison, see if any of the three curves can be approximated by a straight line for the entire length.

(e) Plot on a magnified scale the squares of the numbers between 0 and 0.2, and also between 0 and 0.02. Can these graphs be approximated by straight lines?

6. In 1973 the postal rates for first-class letters were 8 cents for 0 to 1.0 oz, 16 cents for 1.0 to 2.0 oz, 24 cents for 2.0 to 3.0 oz, etc. Plot a graph of these pairs of numbers from zero to 5.0 oz.

7. In an experiment, a coin was tossed 300 times and the frequency of occurrence of runs of different length of successive heads or of successive tails was recorded. The results are plotted in Fig. 6.14.

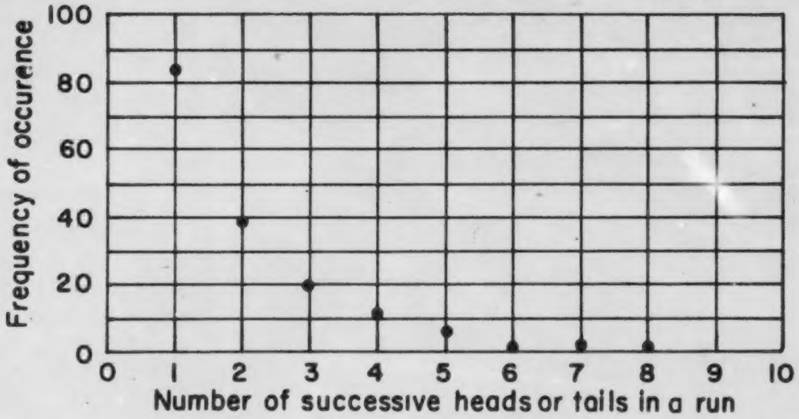


Fig. 6.14

This is an example of a graph where it makes no sense to connect the points or interpolate between them. Each variable can only take on integral values, so that saying that a run of 3.5 heads or tails occurred about 16 times is meaningless.

(a) If the number of occurrences of runs of three successive heads or tails had not been recorded, how would you estimate it?

(b) The graph in Fig. 6.2 has the population of the United States as the dependent variable. This variable, obviously, can have only integral values. Why is each line connecting the points an unbroken line and not a series of points representing integral values?

8. Suppose you knew only the points shown on the graph in Fig. 6.15. At what additional values of x would you like to know the value of y before sketching the graph? Explain.

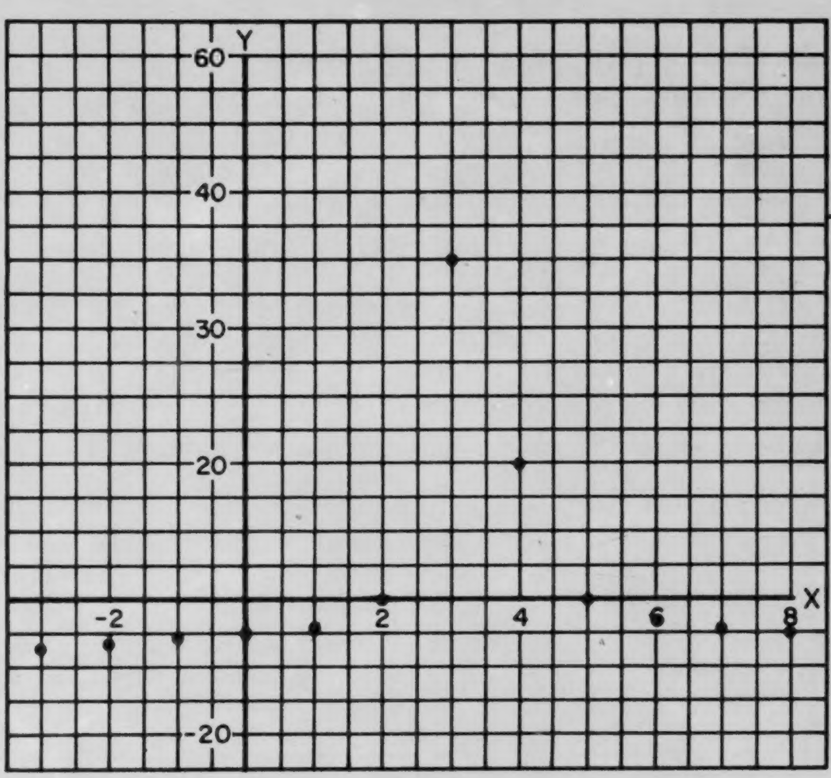


Fig. 6.15

Chapter 7. LINEAR AND POWER FUNCTIONS

7.1 Notation

In the preceding chapter we discussed ways of constructing graphs to display a function. We also pointed out that a graph is not the only way of displaying a function and that, in fact, tables and rules stated verbally or algebraically may also be used.

Whenever we can express a function in algebraic terms, we shall do so for compactness and ease of handling. There are no strict rules on what letters to choose for what purpose, but there are some general conventions which are worth following since they reduce the need for frequent reminders of the meaning of symbols.

Suppose you want to express the rule "to find the value of the dependent variable, square the value of the independent variable and multiply it by some constant." If instead of making this lengthy statement you simply write $a = bc^2$, without any further explanations, you are not guaranteed that it will be correctly interpreted, the reader may actually understand it as "to find the value of the dependent variable $\underline{a}$ take the independent variable $\underline{b}$ and multiply it by some constant $\underline{c}$ squared."

To minimize such misunderstandings the following conventions are useful.

(a) Numbers which are not specified but are meant to have a fixed value for a given function are called parameters and are often expressed by the first letters of the alphabet: $\underline{a}$, $\underline{b}$, $\underline{c}$, $\underline{d}$, ...

(b) Continuous variables are usually expressed by the last letters of the alphabet, such as $\underline{r}$, $\underline{s}$, $\underline{t}$, $\underline{u}$, $\underline{v}$, $\underline{w}$, $\underline{x}$, $\underline{y}$, and $\underline{z}$. In case of angles Greek letters such as α, θ, and ϕ are also used.

(c) Variables which are limited to non-negative integers are usually expressed by the letters $\underline{i}$, $\underline{j}$, $\underline{k}$, $\underline{l}$, $\underline{m}$, and $\underline{n}$. Applying these conventions to the rule which we have just spelled out in words could yield any of the following "spellings"

$$y = ax^2, \ s = bt^2, \ x = cz^2$$

and the chances of misreading this to have the second meaning, $y = b^2x$, are very small.

(d) Often we wish to pick out a number of specific values of a variable. These specific values need not be integers, but they can be labeled x_0, x_1, x_2, or in general x_i or x_n, with the corresponding values for the dependent variable y_0, y_1, y_2, or in general y_i or y_n. The integers serve only to distinguish values of a variable one from another, just as a route number on a bus serves only to identify it. Integers used in this way are called indices.

(e) Many properties of functions can be discussed without spelling out the detailed mathematical rule. Thus, a notation is needed to indicate that one variable is a function of another. The most common one is a shorthand form of the statement "$\underline{y}$ is a function of $\underline{x}$" and is written as $y = f(x)$ read "$\underline{y}$ equals $\underline{f}$ of $\underline{x}$." The notation $f(\ )$ stands for a definite rule relating the dependent variable to the independent variable. The independent variable which is placed in the parentheses in $f(\ )$ is also called the argument of the function. If we write $y = f(x)$ and $u = f(v)$ we call the variables by different names, but the understanding is that the same rule relates $\underline{y}$ to $\underline{x}$ and $\underline{u}$ to $\underline{v}$. If we wish to indicate different rules, we use different letters such as $y = g(x)$, $y = F(x)$, or we use indices such as $x = f_1(t)$, $x = f_2(t)$. etc.

We can think of an equation such as $f(x) = 2x + 3$ as an alternative notation for $y = 2x + 3$. The power of the $f(x)$ notation lies in the ease with which a value of the dependent variable can be specified for a given value of the independent variable; for example, the notation $f(3)$ is used to represent the value that the dependent variable assumes when the independent variable is equal to 3. If $f(x) = 2x + 3$, then $f(3) = 2 \cdot 3 + 3 = 9$. That is,

$f(3)$ is obtained by performing the same sequence of operations upon 3 that are performed upon x in the rule defining the function.

The argument of a function, the entity placed in the parentheses in $f(\ )$, may sometimes not be identical with the independent variable. For example, if $y = f(x-2)$, then $x - 2$ is the argument of the function but x is the independent variable. To find y for a given value of x we first calculate $x - 2$ and then apply the rule $f(\ )$ to $x - 2$. For example, when $x = 7$, $y = f(7-2) = f(5)$.

Questions

1. Given the function $f(x) = 3x + 1$, find (a) $f(10)$, (b) $f(3)$, (c) $f(-1)$.
2. For $f(x) = 5$, find
   - (a) $f(0)$
   - (b) $f(2)$
   - (c) What is the range of this function?
3. Make up a function f such that $f(3) = 7$. Can you make up another function g such that $g(3) = 7$?
4. Let $p = f(s)$ be the perimeter of a square expressed as a function of the length of its side s. What is the rule for $f(s)$? Express in words the meaning of $f(3)$.
5. Let $f(\ )$ stand for the rule "take the square of the argument."
   - (a) What is $f(x-1)$?
   - (b) If $y = f(x-1)$, what is the value of y when $x = 4, 1, -1, -9$?
6. Suppose you have a function $y = f(x)$ such that $f(3) = 10$. If $z = f(x-2)$, for which value of x will $z = 10$?

7.2 Homomorphic Curves

Suppose a certain curve is the graphical display of a function $y = f(x)$. Suppose further that we have another curve which has the same size, shape, and orientation as the first curve (Fig. 7.1). That is, we can conceive of the second curve as being generated from the first by displacing each point of the original curve a fixed amount vertically, and a fixed amount horizontally.

For example, in Fig. 7.1 the dashed curve could have been obtained by displacing the solid curve four units vertically upward and three units horizontally to the left. Whenever two curves are related in this way, we say that the curves are homomorphic.

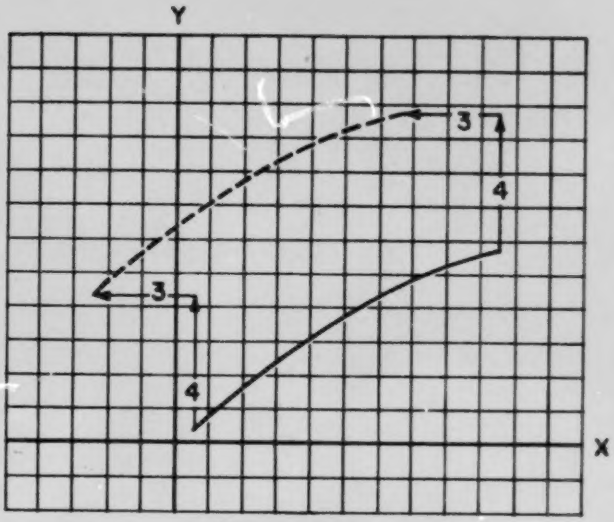


Fig. 7.1

Let the functions describing two homomorphic curves be given by y = f(x) and y = g(x) respectively. How are the two rules f(x) and g(x) related to each other? To find out, it is best to consider the two possible displacements in the plane separately. First, we take a curve displaced only vertically (Fig. 7.2). Since for each value of x the value of y on the

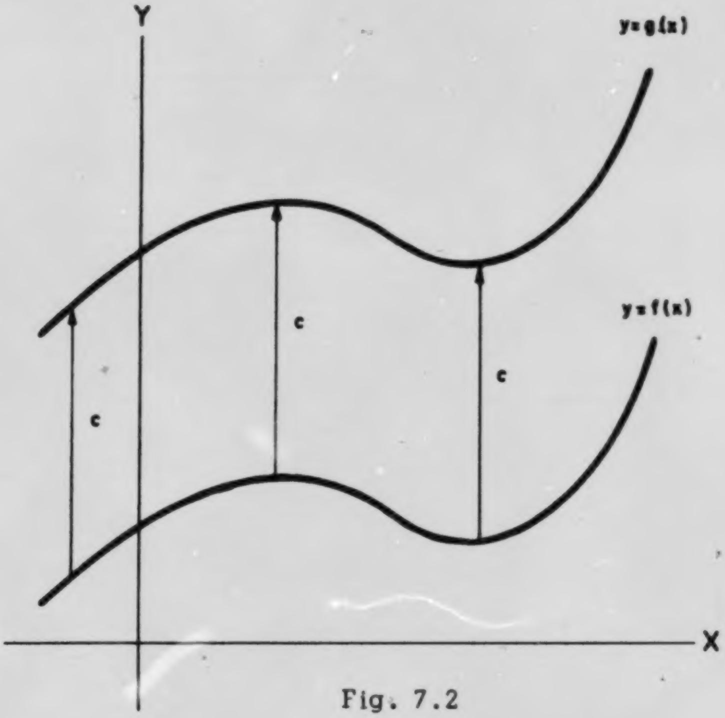


Fig. 7.2

new curve is c units greater than the value of y on the original curve, we have $g(x) = f(x) + c$, where $c > 0$. Thus, in words, to find a y value of the new curve for a given x we use the rule of the original curve and add a number c, indicating the vertical displacement. We can therefore write the rule for the new curve as

$$y = f(x) + c \qquad (1)$$

or

$$y - c = f(x)$$

If the homomorphic curve $y = g(x)$ is c units below the original curve $y = f(x)$ then, by an argument similar to the above, the relationship between the two functions can be expressed as $g(x) = f(x) - c$, where again $c > 0$. We can express both upward and downward displacements by writing only $g(x) = f(x) + c$ and letting c have either positive or negative values.

Let us now take the case of the horizontal displacement shown in Fig. 7.3. The original function is expressed as $y = f(x)$ and the function homomorphic to it as $y = h(x)$, displaced three units horizontally to the right. Consider a given point on the curve $y = h(x)$, say, $x = 7$, for which $h(7) = 5$. For which value of x is $f(x) = 5$? Since the curve corresponding to $y = h(x)$ was generated by displacing the curve corresponding to $y = f(x)$, three units to the right, the value of x, for which $f(x) = 5$, will be three units to the left of 7, i.e., at $x = 4$.

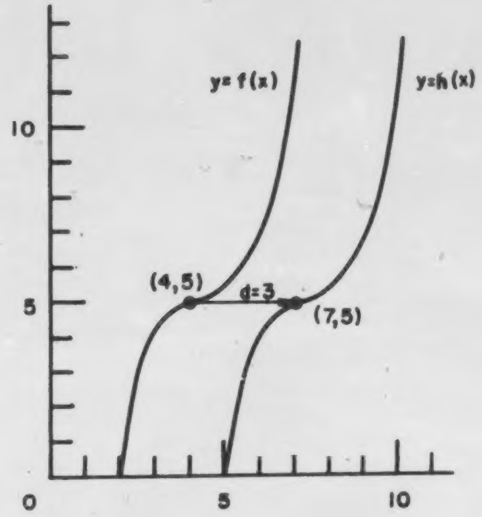


Fig. 7.3

In general, if the curve $y = h(x)$ is generated by displacing the curve $y = f(x)$ to the right by d units ($d > 0$) (Fig. 7.4), then applying the rule h( ) to any value of x will yield the same number as applying the rule f( ) to $x - d$. Hence $h(x) = f(x - d)$ and the rule for the new curve becomes

$$y = f(x - d) \qquad (2)$$

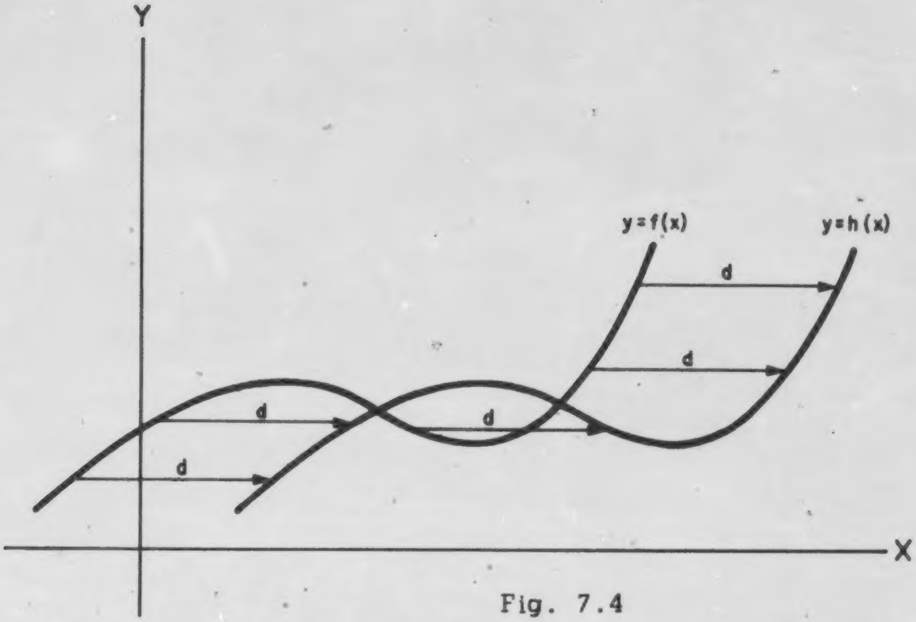


Fig. 7.4

Here again, as with the vertical displacement, our illustration is based on the homomorphic curve being to the right of the original curve. However, just as with downward vertical displacements, by denoting displacements to the left by negative values of d we can use the equation $y = f(x - d)$ for both types of horizontal displacements.

We can now combine these displacements. Given a function $y = f(x)$, the y values on a curve homomorphic to it, displaced c units vertically and d units horizontally is

$$y = f(x - d) + c \qquad (3)$$

or

$$y - c = f(x - d)$$

Questions

1. Given some arbitrary curve, how many curves can be constructed which are homomorphic to it?
2. By use of vertical and horizontal displacements, find whether the curves given in Fig. 7.5 are homomorphic.

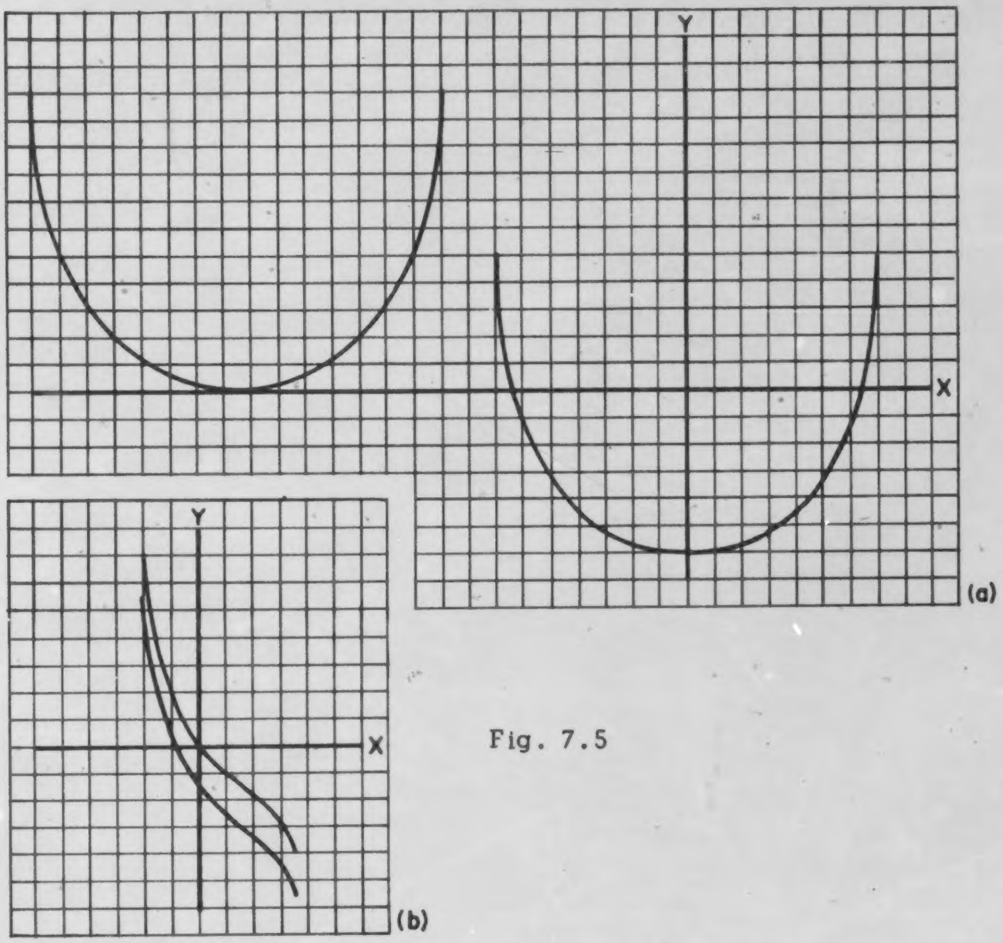


Fig. 7.5

3. If a curve AB is homomorphic to a curve CD, and CD in turn is homomorphic to EF, is AB homomorphic to EF? Give your reasons.

4. In plane geometry one uses the congruence relation: Two plane figures are said to be congruent if one can be exactly superimposed on the other. How is congruence different from homomorphism?

5. A curve is described by the rule $y = x^2$. What is the rule for the curve homomorphic to it, displaced three units horizontally and four units vertically?

6. Each of the functions $f(x)$ below can be displayed by a curve $y = f(x)$. For each case state the functions for the curves displaced by the amounts indicated.

	Horizontal Displacement	Vertical Displacement
(a) $f(x) = 1/(x+1)$	-2	10
(b) $f(x) = 2$	3	-1
(c) $f(x) = x/(x+1)$	1	0

7. Let $\underline{n}$ be the ordinal number of the throw of a die and ℓ the value showing on the top face, ℓ is a function of $\underline{n}$. Call it $g(n)$.

(a) What is the domain and range of $g(n)$?

(b) What is the domain and range of $h(n) = g(n+3) + 10$?

(c) How would a graphical display of $h(n)$ be related to the graphical display of $g(n)$?

7.3 Direct Proportions

A very common relation between a dependent and independent variable is that when the independent variable is doubled or tripled so is the dependent variable. In other words, their ratio is a constant

$$\frac{y}{x} = a$$

Written in the form $y = f(x)$ this says that

$$f(x) = ax \qquad (4)$$

This function is referred to as a direct proportion and the parameter $\underline{a}$ is called the constant of proportionality.

Mathematically the domain and range for this function extends over the entire number line. However, if the variables are not pure numbers but are measures for definite quantities, practical consideration may restrict the domain. For example, the circumference of a circle as a function of its diameter is given by $f(x) = \pi x$. The diameter of a circle cannot be negative, hence in this case the function makes sense only for $x \geq 0$.

Figure 7.6 shows the graphs of several direct proportions for various values of the constant of proportionality. Note that graphs of direct proportions constitute a family of straight lines through the origin each of which can be generated by varying the constant of proportionality.

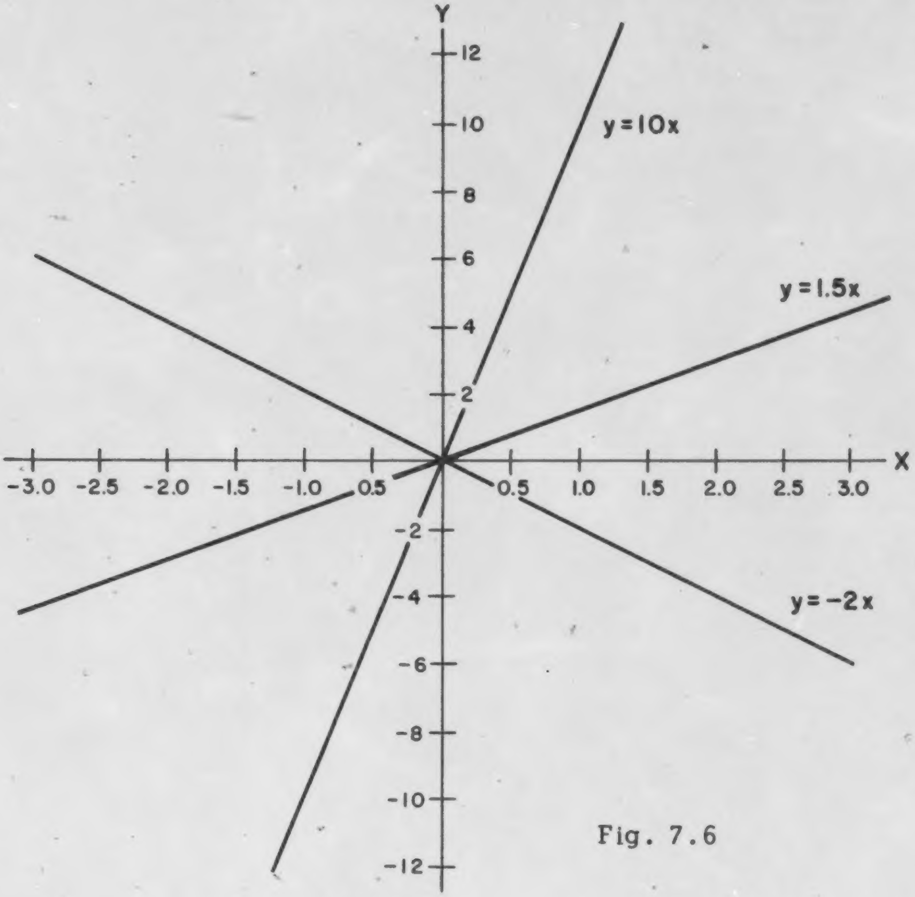


Fig. 7.6

Lines representing functions with positive constants of proportionality are directed upward to the right, while lines which are graphs of functions with negative constants of proportionality are directed downward to the right.

The graphs in Fig. 7.6 could represent a multitude of real situations. For example, the two lines with a positive constant of proportionality, could represent the mass of a liquid as a function of its volume (for $x \geq 0$) or the position of a point on a line as a function of time. The line with the negative constant of proportionality might represent the force exerted by a spring as a function of its stretch; the negative value indicates that the force is opposite in direction to the stretch.

The steepness of the line is related to the value of the proportionality constant. Functions having proportionality constants whose absolute value is larger have steeper graphs than functions with constants of proportionality whose absolute value is smaller. To obtain the constant of proportionality from the graph of the function we can choose any point on the graph and divide it's y-coordinate by its x-coordinate. A note of caution is needed here: the values of the corresponding coordinates must be obtained by reading them off the scales used along each axis. These scales, on the two axes may be different as in Fig. 7.6, thus finding the values of the coordinates by measuring along both axes with a ruler would result in errors.

When the dependent and independent variables have different units, then the constant of proportionality has the units of a specific quantity arrived at by dividing the unit of the dependent variable by the unit of the independent variable (see Section 2.4). Thus in the examples which we have just mentioned, the constant of proportionality defined the following specific quantities respectively: density (mass per unit volume), velocity (displacement per unit time) and the force constant (force per unit length).

Questions

1. The relation between feet and yards is given by the equation $y = 3x$.
   (a) What does $\underline{y}$ represent? What does $\underline{x}$ represent?
   (b) Interpret 3, the constant of proportionality.
2. Which of the following functions are approximately direct proportions? For the cases which are, write the corresponding equation and indicate reasonable values for the domain and range of the function for which you expect the direct proportion to be valid.
   (a) The height of a building and the number of floors.
   (b) Age and weight.
   (c) Weight of a package and price of postage.
   (d) Number of telephone calls and telephone bill.
   (e) Weight of patient and amount of medication.
   (f) Age of tree and thickness of tree.

3. For a certain kind of paper one sheet is 3×10^{-3} cm thick;

(a) Plot the thickness of a book made with this paper as a function of the number of pages it has.

(b) What is the constant of proportionality?

(c) What is the algebraic formula relating the number of pages and thickness?

4. Figure 7.7 represents the masses of samples of some substance and the corresponding volumes.

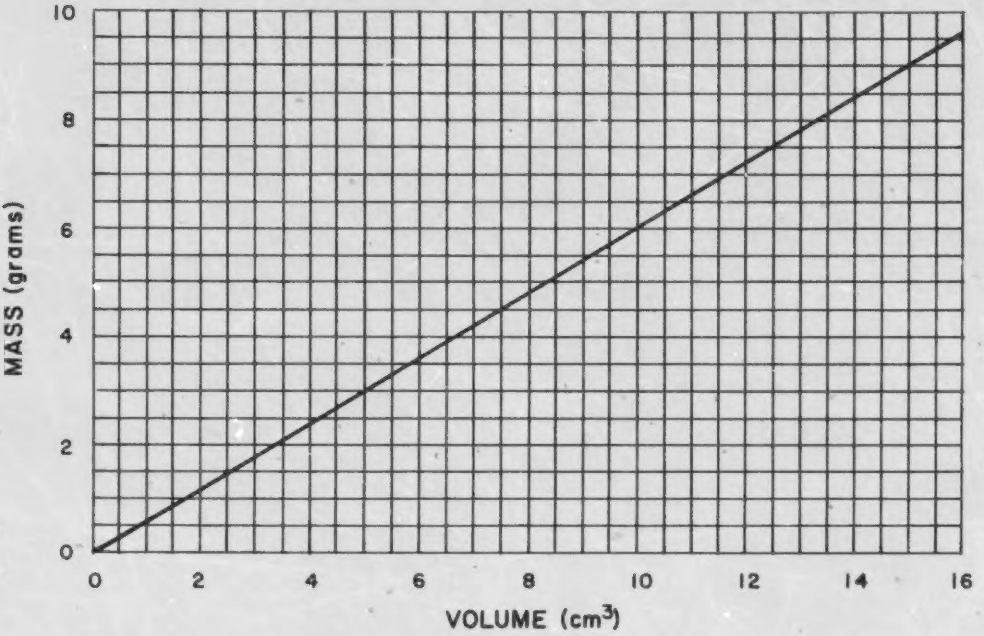


Fig. 7.7

(a) What is the mass per cm^3 of this substance?

(b) What is the function that corresponds to this straight line?

(c) What is the constant of proportionality?

(d) On the same graph draw the line corresponding to mass vs. volume of water.

(e) Given two lines on a mass-volume graph, how can one readily see which corresponds to the denser substance?

5. What is the equation of the straight line through the origin and the point (-3,6)? Through the point (100,1)?

6. A straight line passes through the origin and the point (c,d). How can one tell the sign of the constant of proportionality of the corresponding relation by just looking at the signs of c and d?

7. Often a constant of proportionality is given as a rate. What proportion is implied by the statements:
   (a) the rate of exchange of Swiss francs is 0.32 dollar per franc?
   (b) the rate of flow of water over a dam is 50 cm^3 per minute?
   (c) the rate of interest is 10 per cent?

8. An electronics firm lists the following prices for different quantities of a certain brand of capacitors as follows:

   Lots of 1 - 24 at 48¢ each
   Lots of 25 - 49 at 35¢ each
   Lots of 50 and up 27¢ each

   (a) Plot the cost of the capacitors as a function of their number.
   (b) Would you order 23 capacitors?

7.4 The Linear Function

We have seen that the relation y = ax describes a whole family of graphs, straight lines passing through the origin, whose steepness is determined by the value of a. We can, of course, construct lines that do not pass through the origin and are homomorphic to a line described by y = ax, by displacing each point on the line corresponding to y = ax by a fixed amount b in the vertical direction. This procedure changes any function y = f(x) into y = f(x) + b (Section 7.2). Therefore the function described in the graph homomorphic to the graph of y = ax becomes

$$y = ax + b \tag{5}$$

Because the graph described by this equation is a straight line the function f(x) = ax + b is called a linear function. This function has two parameters a ($a \neq 0$), and b. Note that for x = 0, y = b. Thus the line crosses the y-axis at y = b. For this reason the parameter b is called the y intercept. Figure 7.8 shows several lines homomorphic to the line given y = 0.5x, which were obtained by varying b, the y intercept.

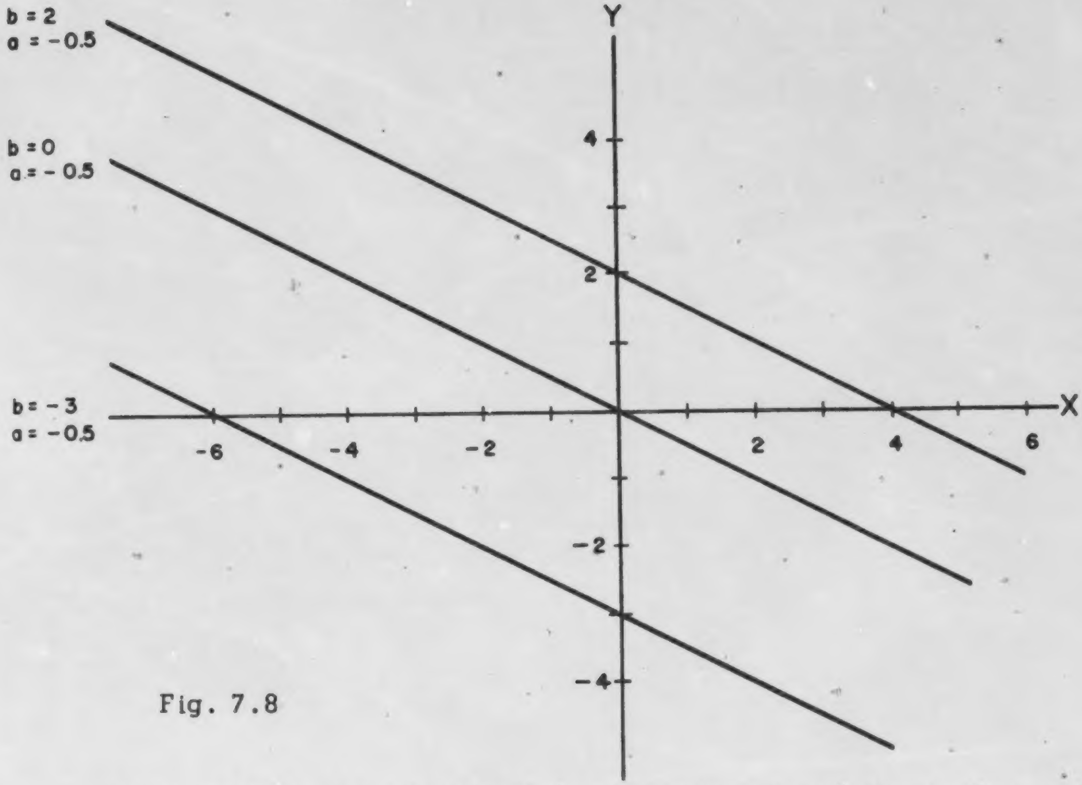


Fig. 7.8

To investigate the meaning of the parameter a in relation to the graph of a linear function, we graph some functions with the same value of b but different values of a (Fig. 7.9). As was the case with the direct proportion,

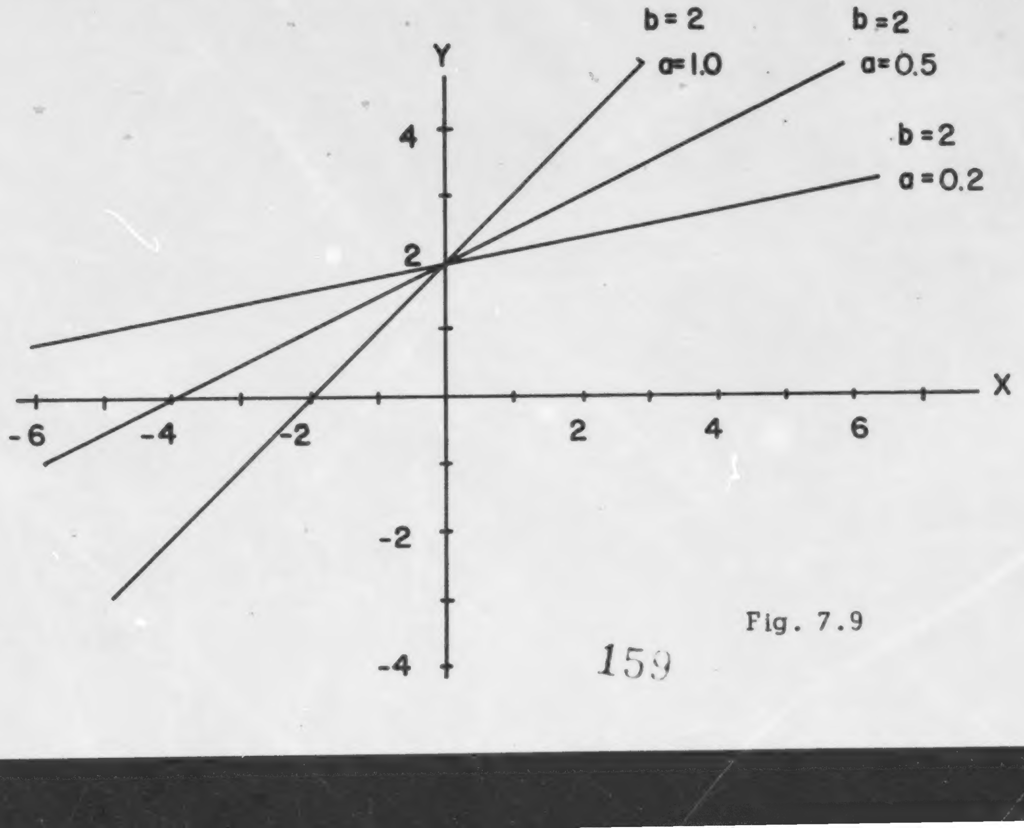


Fig. 7.9

the steepness of the line is determined by a which is therefore called the slope of the linear function.

If a = 0, Equation (5) becomes

$$y = b$$

that is, the dependent variable has the same value for all values of the independent variable. For this reason, y = b is sometimes referred to as a constant function. An example of such a function is the graph of the density of pieces of aluminum versus their volumes (Fig. 7.10). The constant function is not considered a special case of the linear function since a linear function by definition has a first degree term in its independent variable.

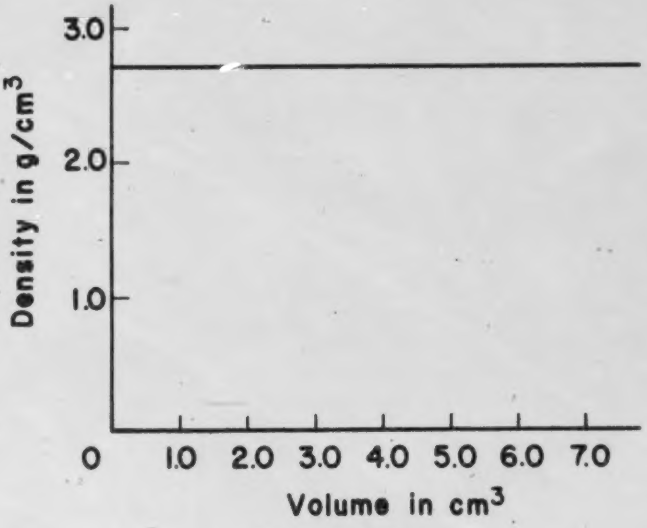


Fig. 7.10

Questions

1. Lines homomorphic to that given by y = ax can also be generated by moving each point a given amount horizontally: y = a(x - d). Does this procedure yield any straight lines that cannot be generated by a vertical displacement of the form y = ax + b?

2. Under what conditions does the function y - c = a(x - d) describe the same straight line as y = ax?

3. The centigrade and Fahrenheit temperature scales are related as follows: 0°C is equal to 32°F, and each 1°C is equal to 1.8°F.

   (a) Express the temperature in degrees Fahrenheit as a function of the temperature in degrees centigrade.

   (b) Plot the corresponding graph.

   (c) Is there a temperature which is expressed by the same number on both temperature scales?

4. The following table was taken from a Federal Income Tax Brochure:

Taxable Income		Tax	
Not over $500		14% of the Amount	
Over -	But not Over -		of Excess Over -
$500	$1,000	$ 70+15%	$500
$1,000	$1,500	$145+16%	$1,000
$1,500	$2,000	$225+17%	$1,500
$2,000	$4,000	$310+19%	$2,000

   What kind of function describes the dependence of the tax on the taxable income? Plot the corresponding graph.

5. Write a computer program to calculate the income tax for taxable incomes up to $4,000. Use the information given in the preceding problem.

6. A straight line parallel to the y axis does not describe a linear function. Why?

7. The equation of a straight line can also be written in the form $\frac{x}{m} + \frac{y}{n} = 1$. What are the geometric meanings of m and n?

7.5 Finding the Equation of a Straight Line

Often the purpose of a scientific experiment is to determine if there exists a simple functional relationship between two quantities — for example, between the volume of a gas and its temperature. To get a feel for the nature of the relationship, the experimental data are usually graphed. As you become familiar with the graphs of some simple fundamental functions, you will often be able to get an idea of what kind of functional relationship might exist between the two quantities by looking at the graph.

A linear relationship is the easiest to recognize because it is only necessary to decide if the points representing the experimental data (taking into consideration the uncertainties of the measurements) lie close to a straight line (Fig. 7.11(a)). If they do, then you have to decide how to draw the line best fitting the points (Fig. 7.11(b)). To do this reasonably well requires practice, which is best obtained by actually doing experiments in a laboratory and graphing the data. However, once the line is drawn, to find its equation, that is, to determine the values of the parameters a and b in the expression

$$y = ax + b$$

is a purely mathematical question.

Suppose we end up with a straight line like the one in Fig. 7.12. We choose two points, P and Q on the line (as far apart as possible), and note their respective coordinates (x_1, y_1) and (x_2, y_2). Since the points P and Q are on the straight line their coordinates satisfy the equation

$$y_1 = ax_1 + b$$

and

$$y_2 = ax_2 + b$$

We can solve these two equations for a and b in two steps. First we subtract the first equation from the second and get

$$y_2 - y_1 = a(x_2 - x_1)$$

or

$$a = \frac{y_2 - y_1}{x_2 - x_1} \qquad (6)$$

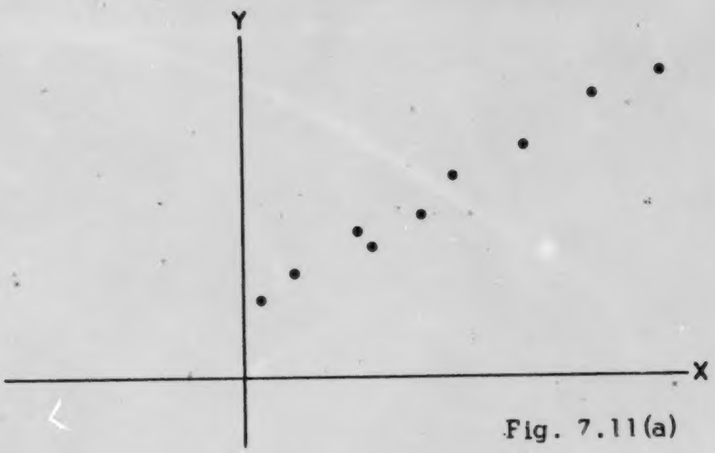


Fig. 7.11(a)

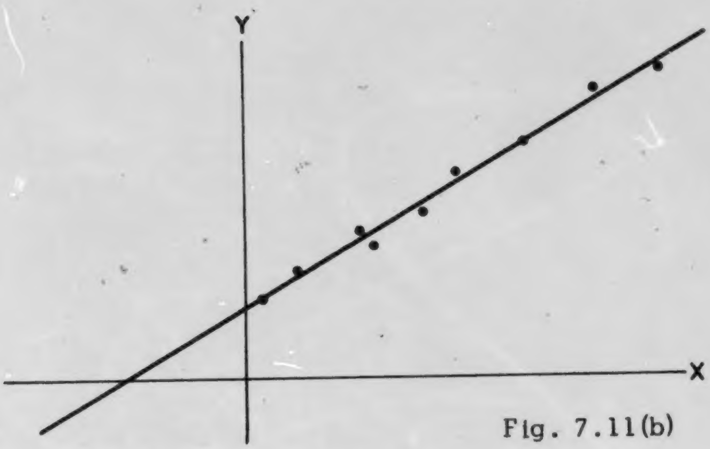


Fig. 7.11(b)

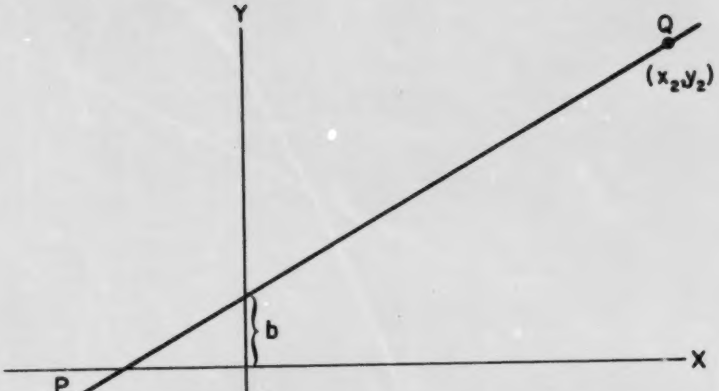

Y
Q
(x_2,y_2)
b
X
P
(x_1,y_1)

Fig. 7.12

Using the now known value of a we can calculate b from

$$b = y_1 - ax_1$$

If the slope a of a straight line passing through the point (x_1, y_1) is given, then we can find the equation of the line simply by substituting the value of b from the last equation in Equation (5):

$$y = ax + y_1 - ax_1$$

or

$$y = y_1 + a(x - x_1) \tag{7}$$

In the preceding section the parameter a was defined as the slope. As Equation (6) shows, it can be obtained by dividing the change in the dependent variable by the corresponding change in the independent variable. Thus, the slope gives the rate of change of the dependent variable with respect to the independent variable. Any two points on a given straight line yield the same slope. Therefore, the slope of a straight line is a property of the whole line. We shall see in the next chapter that curved lines do not have this property.

Since only the change in coordinates enters into the calculation of a slope Equation (6) is often written as

$$a = \frac{\Delta y}{\Delta x}$$

Questions

1. A straight line passes through the points (-2, 3) and (-4, 4).
   (a) Draw the line.
   (b) What are the values of Δx and Δy.
   (c) Use them to find the slope and the y-intercept of the line.
2. A straight line passes through the points with the coordinates (x_1, y_1) and (x_2, y_2).
   (a) Under what conditions will the slope be positive? Negative?
   (b) How would you describe in words a line with negative slope.

3. Suppose you plot the displacement of a moving point as a function of time and find that the points fit a straight line of the form $y = ax + b$, where y is given in meters and x is given in seconds. What are the units of a and b?

4. Find the equation of each of the lines in Fig. 7.13.

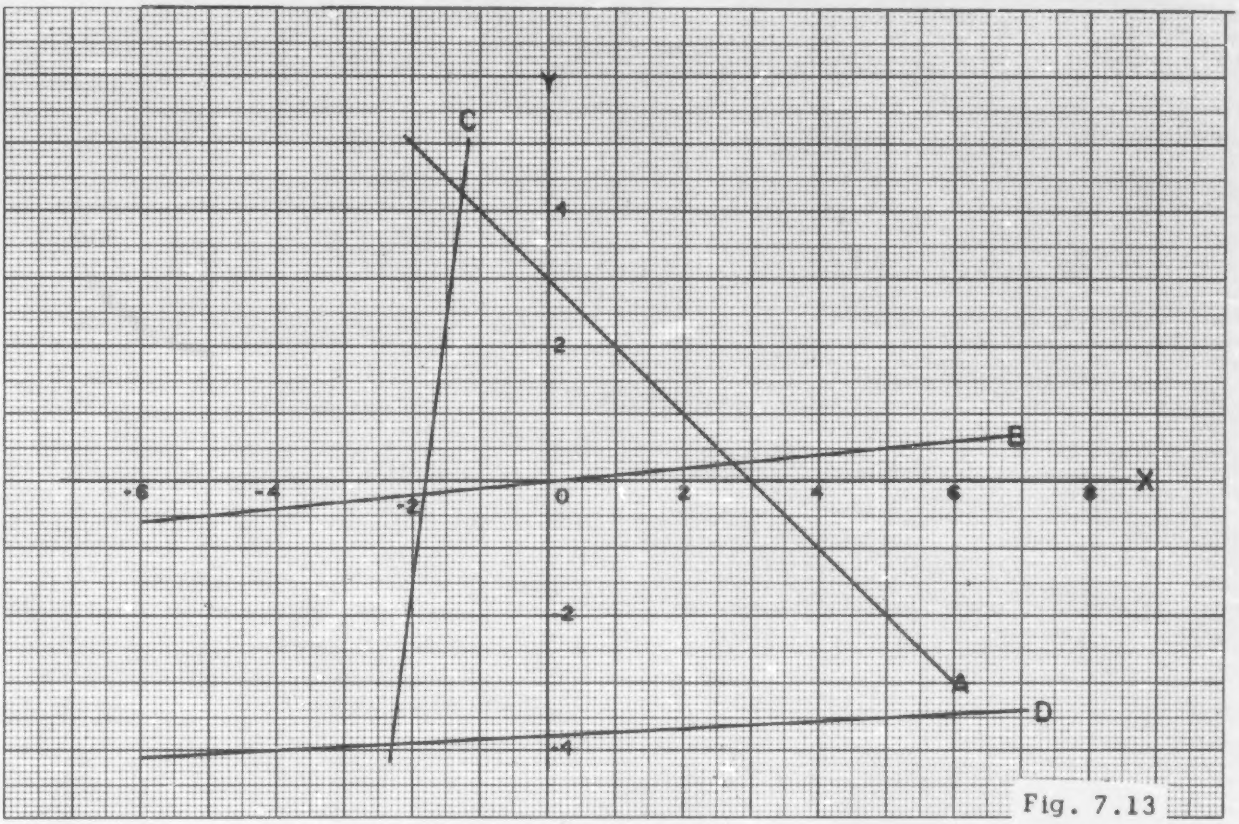


Fig. 7.13

5. (a) What is the equation of the line through (-1, 2) with slope -5?

   (b) What is the equation of the line through (5, 6) with slope $\frac{1}{4}$?

6. Write a computer program which computes the values of the parameters a and b given the coordinates of any two points on the line.

7. The slope of a straight line is independent of the two points selected to calculate it. Yet the text suggests to choose these points as far apart as possible. Why?

8. Which of the tables below describe a linear function?

Hint: The slope of the linear function $(y_2 - y_1)/(x_2 - x_1)$ is a characteristic of the function and is the same for any values of x_1 and x_2 which are used to calculate it.

(1)

x	y
-1	-3
0	-5
1	-7
2	-9
3	-11
4	-13
5	-15

(2)

x	y
0.1	2.01
0.2	3.01
0.3	4.01
0.4	5.01
0.5	6.01
0.6	7.01
0.7	8.01

(3)

x	y
1	2
2	4
3	8
4	16
5	32
6	64

7.6 The Quadratic Function

In this section we shall study functions of the form $y = ax^2$, where a is a constant. The curve corresponding to this function is called a parabola. Figure 7.14 shows a number of parabolas corresponding to different values of a. As you can see, the parabolas corresponding to positive values of a have their branches pointing upward, whereas parabolas corresponding to negative values of a have their branches pointing downward. Notice that all the graphs are symmetrical about the y axis; that is, the points on the curves to the right of the y axis will fall directly on the corresponding points to the left of the y axis if the graph paper is folded along the y axis. The axis of symmetry, the y axis, is called the axis of the parabola, and the point of the parabola which lies on the axis of symmetry is called the vertex of the parabola.

Any curve that is homomorphic to a curve whose equation is $y = ax^2$ is also a parabola. Figure 7.15 shows a parabola whose equation is $y = ax^2$ and another parabola homomorphic to it with the equation

$$\begin{aligned} y - n &= a(x - m)^2 \\ &= ax^2 - 2amx + am^2 \end{aligned}$$

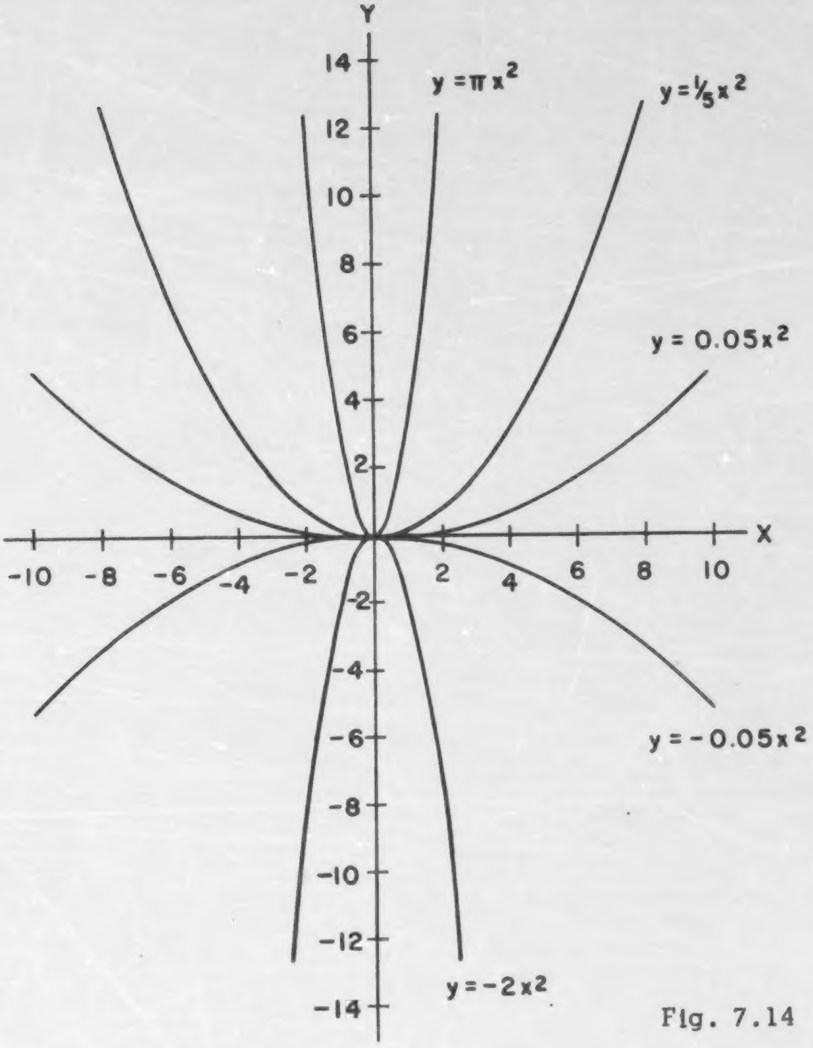


Fig. 7.14

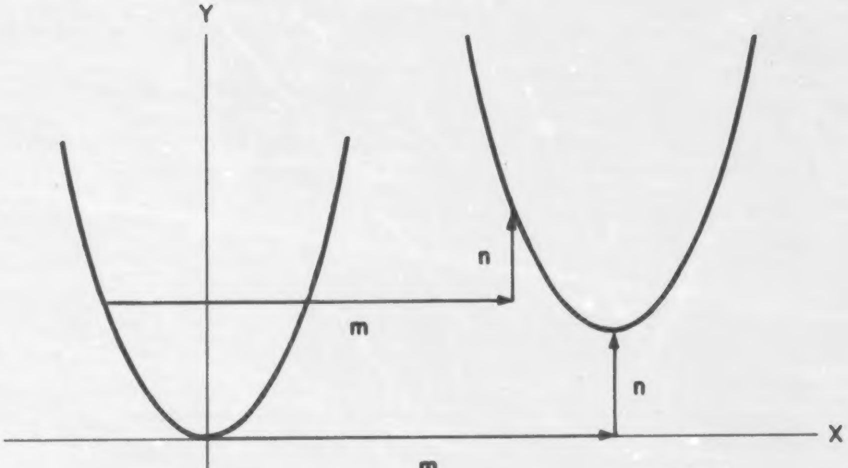


Fig. 7.15

As in the case of the linear function, we can find a standard form for this equation. Because a, m, and n are constants, so are the combinations $-2am$ and $(am^2 + n)$, and we shall call the first combination b and the second combination c; that is, $-2am = b$ and $am^2 + n = c$. Thus, any parabola whose axis is vertical is described by a function of the form

$$y = ax^2 + bx + c$$

where $a \neq 0$. This is called a quadratic function.

Any parabola with a vertical axis is the graph of a quadratic function. The converse is also true: any quadratic function $y = ax^2 + bx + c$ describes a parabola with a vertical axis. This is so because given a, b, and c we can always transform the equation $y = ax^2 + bx + c$ into an equation of the form $y - n = a(x - m)^2$, which describes a parabola with its vertex at (m, n). To do this we solve the two equations

$$-2am = b$$
$$am^2 + n = c$$

for m and n. From the first equation we get $m = -\frac{b}{2a}$. From the second we get $n = (c - am^2)$ and substituting the value of m we have just found, we get, for the two constants (if $a \neq 0$),

$$m = -\frac{b}{2a}$$

and

$$n = c - \frac{b^2}{4a}$$

Using these values of m and n in the equation for a parabola $y - n = a(x - m)^2$, we have

$$y - (c - \frac{b^2}{4a}) = a[x - (-\frac{b}{2a})]^2$$

We have shown, therefore, that any quadratic equation of the form $y = ax^2 + bx + c$ describes a parabola with a vertical axis and a vertex at the point $(-\frac{b}{2a}, c - \frac{b^2}{4a})$.

Questions

1. Consider the function $f(x) = \pi x^2$. What is its domain and range (i) if the variables have no geometric interpretation, and (ii) if they describe the area of a circle as a function of its radius.

2. Sketch the parabolas corresponding to the following functions. Try to guess the general shape of each parabola in the given domains before sketching it. Make each sketch on a different sheet of paper.

   $$\left.\begin{array}{ll}\text{(a)} & y = \frac{3}{2}x^2 \\ \text{(b)} & y = -\frac{3}{2}x^2\end{array}\right\} \quad \text{for } -4 \leq x \leq 4 \quad \text{and} \quad 10 \leq x \leq 15$$

   $$\left.\begin{array}{ll}\text{(c)} & y - 3 = (x - 2)^2 \\ \text{(d)} & y - 1 = (x - 2)^2\end{array}\right\} \quad -1 \leq x \leq 5$$

   (e) $y = x^2 - 8x + 18$ $\qquad 0 \leq x \leq 2$

3. Express the parabola $y - 3 = 2(x-2)^2$ in the form $y = ax^2 + bx + c$.

4. Express the parabola $y = 2x^2 - 4x + 9$ in the form $(y - m) = a(x - n)^2$.

5. Does the equation $y = 2x^2 + 2x$ describe a parabola?

6. Describe the axis of a parabola when the constant $\underline{b}$ in the equation $y = ax^2 + bx + c$ is zero.

7. What is the effect on the parabola of changing the constant $\underline{c}$ in $y = ax^2 + bx + c$?

8. On a piece of graph paper mark off an $\underline{x}$ scale running from -3 to 3. Mark off a $\underline{y}$ scale from -10 to 10. Graph each of the following functions using these scales.

   (a) $x^2 - 50x + 100$.

   (b) $0.05\,x^2 + 2x + 3$

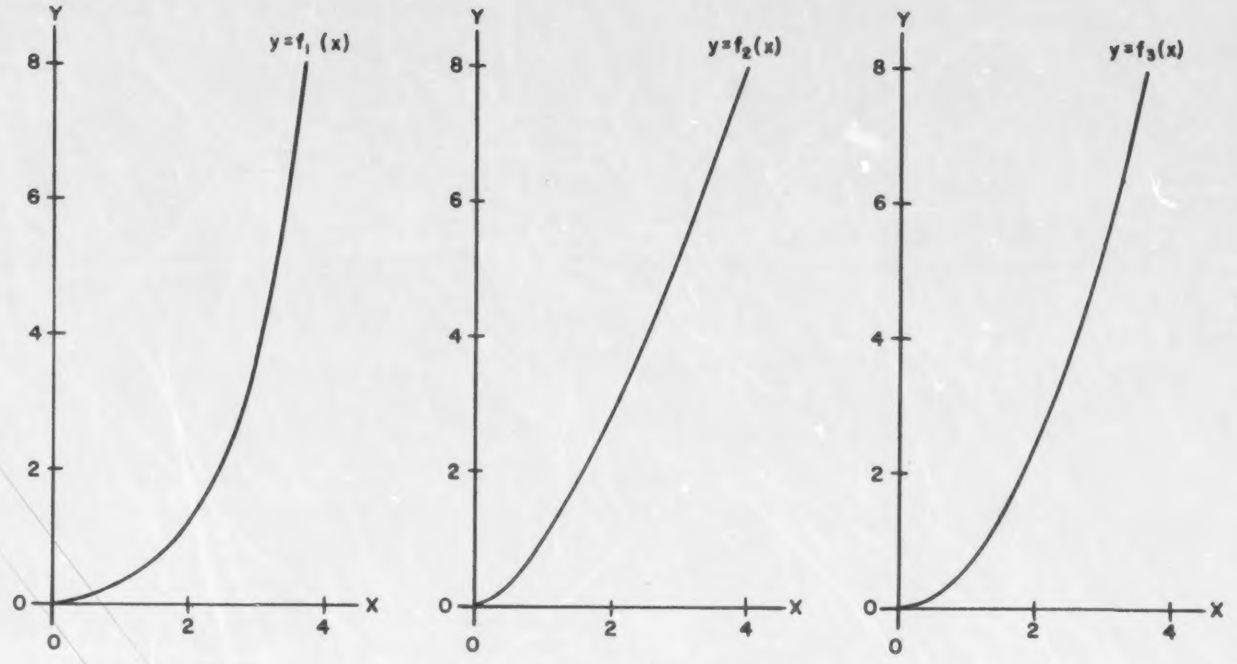
   (c) Give an equation for a parabola which is very nearly a horizontal line for this range of values of $\underline{x}$. Plot the parabola.

7.7 Recognizing Quadratic Functions from Graphs

In Fig. 7.16, three different functions of x, $f_1(x)$, $f_2(x)$, and $f_3(x)$ are graphed. The question is: Do any of the graphs correspond to a function of the type $y = ax^2$, describing a parabola with its vertex at the origin? This is the kind of question which arises when you have a curve passing through the origin with a shape similar to one of the curves in Fig. 7.16.

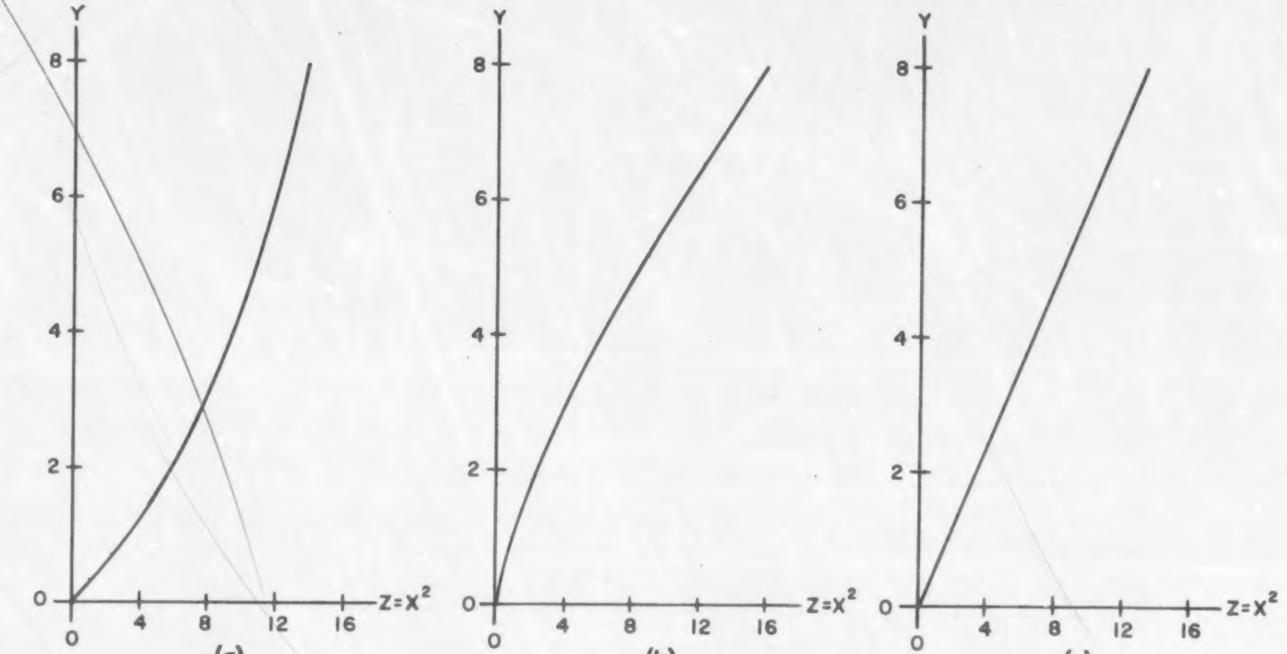
A simple way to answer the question consists of calculating x^2 for a number of values of x and graphing y as a function of a new variable $z = x^2$ instead of as a function of x. If any of the graphs in Fig. 7.16 is, in fact, a graph of the type $y = ax^2$, then the graph of $y = az$ must be a straight line with the slope a. Figure 7.17 shows the results we get when we graph y vs. $z = x^2$ using values for x and y obtained from Fig. 7.17. Only the graph in Fig. 7.17(c) is a straight line with a slope 0.6. We therefore infer that $f_3(x) = 0.6x^2$, whereas $f_1(x)$ and $f_2(x)$ do not express a function of the type $y = ax^2$.

Another way to find out if a set of data is compatible with a relationship of the type $y = ax^2$ is to calculate $\frac{y}{x^2}$ for different points (x, y) on the original curve and see if this fraction remains approximately constant. Table 7.1 gives the result of such a calculation made for values of x and y taken from Fig. 7.16. As you can see, the ratio $\frac{f_3(x)}{x^2}$ is constant within the accuracy to which the graph can be read.



(a) (b) (c)

Fig. 7.16



(a) (b) (c)

Fig. 7.17

TABLE 7.1

x	$\frac{f_1(x)}{x^2}$	$\frac{f_2(x)}{x^2}$	$\frac{f_3(x)}{x^2}$
0.0	--	--	--
0.3	0.78	1.78	0.56
0.5	0.52	1.44	0.60
0.8	0.39	1.13	0.59
1.0	0.34	1.00	0.60
1.5	0.31	0.82	0.60
2.0	0.32	0.71	0.60
2.5	0.36	0.63	0.60
3.0	0.42	0.58	0.60
3.5	0.52	0.54	0.60
4.0	0.67	0.50	0.60

If a graph does not pass through the origin, it cannot, of course, be described by a function of the form $y = ax^2$, but it could possibly correspond to the more general form $y = ax^2 + bx + c$. If the graph clearly indicates a point that is a maximum or a minimum and is symmetrical about a vertical line perpendicular to the curve at this point, it is possible that this point is the vertex (m, n) of a parabola and we can look for a relation of the form $y - n = a(x - m)^2$. We can do this by plotting $y - n$ as a function of $(x - m)^2$ to see if we get a straight line whose slope equals a. Or we can calculate $\frac{y - n}{(x - m)^2}$ to see if this fraction remains very nearly constant for different points on the curve. If so, the fraction is the value of a. We can then write the equation describing the curve, since we now know the values of the three parameters (a, m, and n) needed to specify the particular parabola with which we are dealing.

If it is not possible to determine the position of a possible vertex other methods must be used. We shall illustrate one of these methods which can be used to check whether a graph is a parabola and, if so, to find the particular quadratic function which describes it. Consider Fig. 7.18 (notice

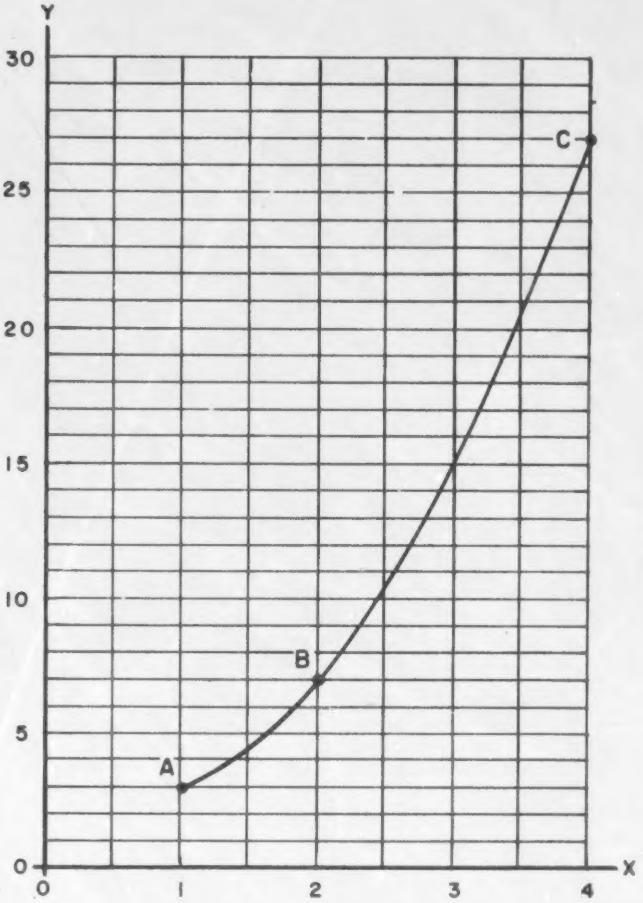


Fig. 7.18

that we have used different scales on the x and y axes to best utilize the graph paper). The problem is to check if the function could be of the form

$$y = ax^2 + bx + c \tag{8}$$

To find the values of the three parameters a, b, and c, we need three equations. The coordinates of any three points on the graph should satisfy Equation (8). Choosing three such points, A, B, and C on the graph (A and C near the end, and B near the center) with coordinates (x_1, y_1), (x_2, y_2), and (x_3, y_3), we have the required three equations:

$$y_1 = ax_1^2 + bx_1 + c$$

$$y_2 = ax_2^2 + bx_2 + c$$

$$y_3 = ax_3^2 + bx_3 + c$$

Here the values of (x_1, y_1), (x_2, y_2), and (x_3, y_3) are known and a, b, and c are unknown. They can be found by solving the three equations for a, b, and c.

For the coordinates of the points A, B, and C in Fig. 7.18, the three equations become

$$\begin{aligned} 3 &= a + b + c \\ 7 &= 4a + 2b + c \\ 27 &= 16a + 4b + c \end{aligned}$$

which have the solution*

$$a = 2, \quad b = -2, \quad \text{and} \quad c = 3$$

Using these values for a, b, and c as coefficients in the equation $y = ax^2 + bx + c$ gives the function

$$y = 2x^2 - 2x + 3$$

describing a parabola that passes through the points A, B, and C on the curve we started with. The question is whether the graph of this function will also pass through (or at least close to) the other points on the original curve. To investigate this we can proceed in different ways. We can draw the graph corresponding to the equation directly on the graph of the original curve or we can determine the position (x_0, y_0) of the vertex of the parabola by using the coordinates for the vertex $(-\frac{b}{2a}, c - \frac{b^2}{4a})$ mentioned in Section 7.6, and see if $\frac{y - y_0}{(x - x_0)^2}$ is nearly constant for points along the original curve. In practice we shall find small deviations, and we shall have to decide whether the deviations, whose size, in part, depends on the magnitude of the errors in the experimental data used in making the original graph, are sufficiently small to allow us to use the equation to describe the experimental data. In general, we cannot expect that the coefficients in an equation derived from a curve made from the experimental data will be represented by small integral numbers, as in our example, so the actual work in calculating a, b, and c will be somewhat harder.

---

*The solution of simultaneous linear equations is discussed in the appendix.

Questions

1. Which of the curves in Fig. 7.19 to 7.21 are parabolas, and what are their corresponding functions?

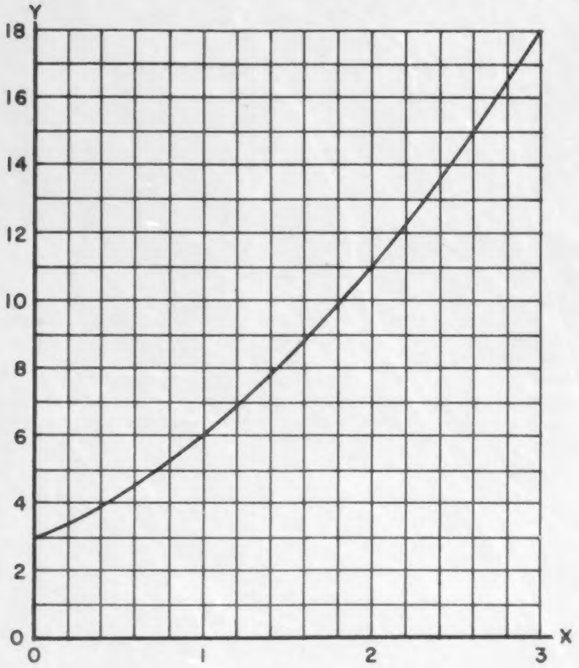


Fig. 7.19

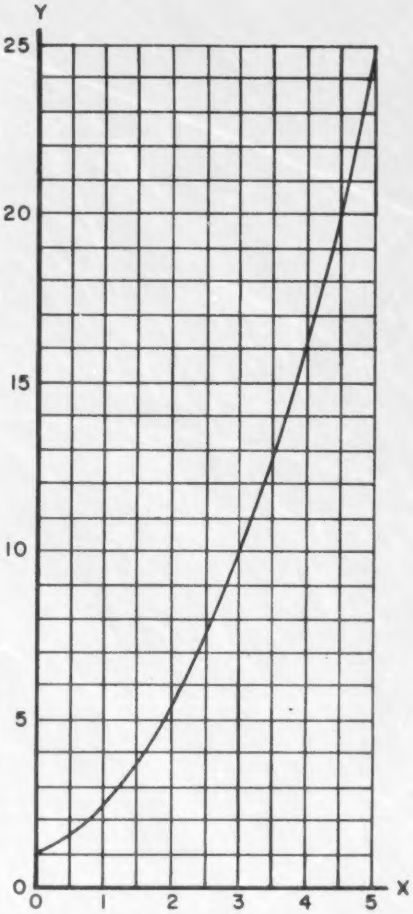


Fig. 7.20

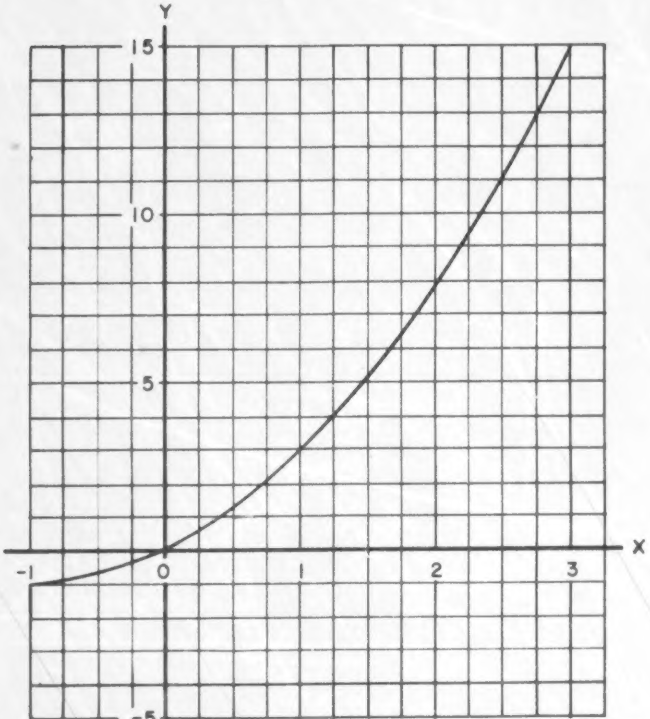


Fig. 7.21

2. What restriction is placed on the parameters of a parabola if
   (a) one insists it go through (0,0)?
   (b) one insists it go through (1,2)?
   (c) one insists it go through both (0,0) and (1,2)?
   (d) one wants it to pass through (0,0), (1,2), and (2,4)?
3. (a) Try to fit a parabola to three collinear points (points lying on the same straight line) by using the procedure of Question 2 on the points (0,0), (1,1), (2,2).
   (b) What do you think happens in general when one tries to fit a parabola to three collinear points?
   (c) At how many points can a straight line intersect a parabola? What relation has this to your answer for (b)?
4. How would you extend the method for recognizing a graph corresponding to $y = ax^2$ to recognize a graph corresponding to $y = ax^3$? $y = ax^n$? Are there any restrictions on the value of n?

7.8 Inverse Proportions

We saw in Section 7.3 that a direct proportion is a relation between two variables such that when the independent variable is doubled so is the dependent variable. Another frequently occurring relation is one where the effect of doubling the independent variable is just the opposite. In other words, when the independent variable is doubled the dependent variable is halved, or in general

$$f(x) = \frac{a}{x} \qquad (9)$$

This function is referred to as an inverse proportion. Note that for any value of $x > 0$ or $x < 0$ the corresponding value of y is uniquely determined, however, when $x = 0$ the function is undefined and we do not have a value for y. Therefore, the domain of the inverse proportion consists of all values of x except $x = 0$. Just as with the direct proportion, however, if the variables are measures of definite quantities, practical considerations may restrict the domain. For instance, the volume of a gas as a function of its pressure is given by

$f(v) = \frac{a}{v}$. Since volume cannot be negative, the function is meaningful only for $v > 0$.

In Fig. 7.22 the graphs of $y = \frac{a}{x}$, each called a rectangular hyperbola, are drawn for $a = 2$ and $a = -2$. Notice that each rectangular hyperbola consists of two branches or parts. When a is positive, the branches lie in the first and third quadrants of the coordinate system (Fig. 7.22(a)), however, when a is negative they lie in the second and fourth quadrants (Fig. 7.22(b)). Furthermore, for $a > 0$ the line $y = -x$ is the axis of symmetry of the whole figure while the line $y = x$ is the axis of symmetry of the individual branches. For $a < 0$ the two axes of symmetry are interchanged.

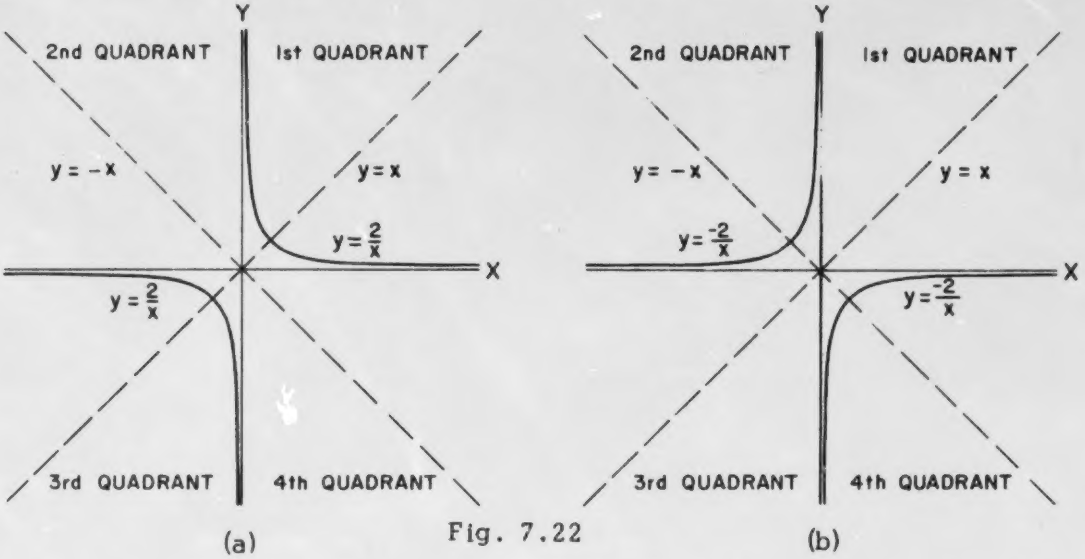


Fig. 7.22

The family of rectangular hyperbolas described by the function $y = \frac{a}{x}$ in Fig. 7.23 shows how the value of a affects the rectangular hyperbolas.

In all cases, note that as x increases, y decreases and the curve gets closer to the x axis, but never meets it. This can be seen by considering the analytic expression $y = \frac{a}{x}$ describing the curve. If we choose $x >> a$ (read "x much greater than a") y becomes much smaller than 1 and the curve is close to the x axis. But we can always choose a still larger value for x, making y even smaller but still not zero. In fact, no matter how large we make x, the value of y will never be zero and the curve will never meet the x axis. The same behavior occurs for very large negative values of x.

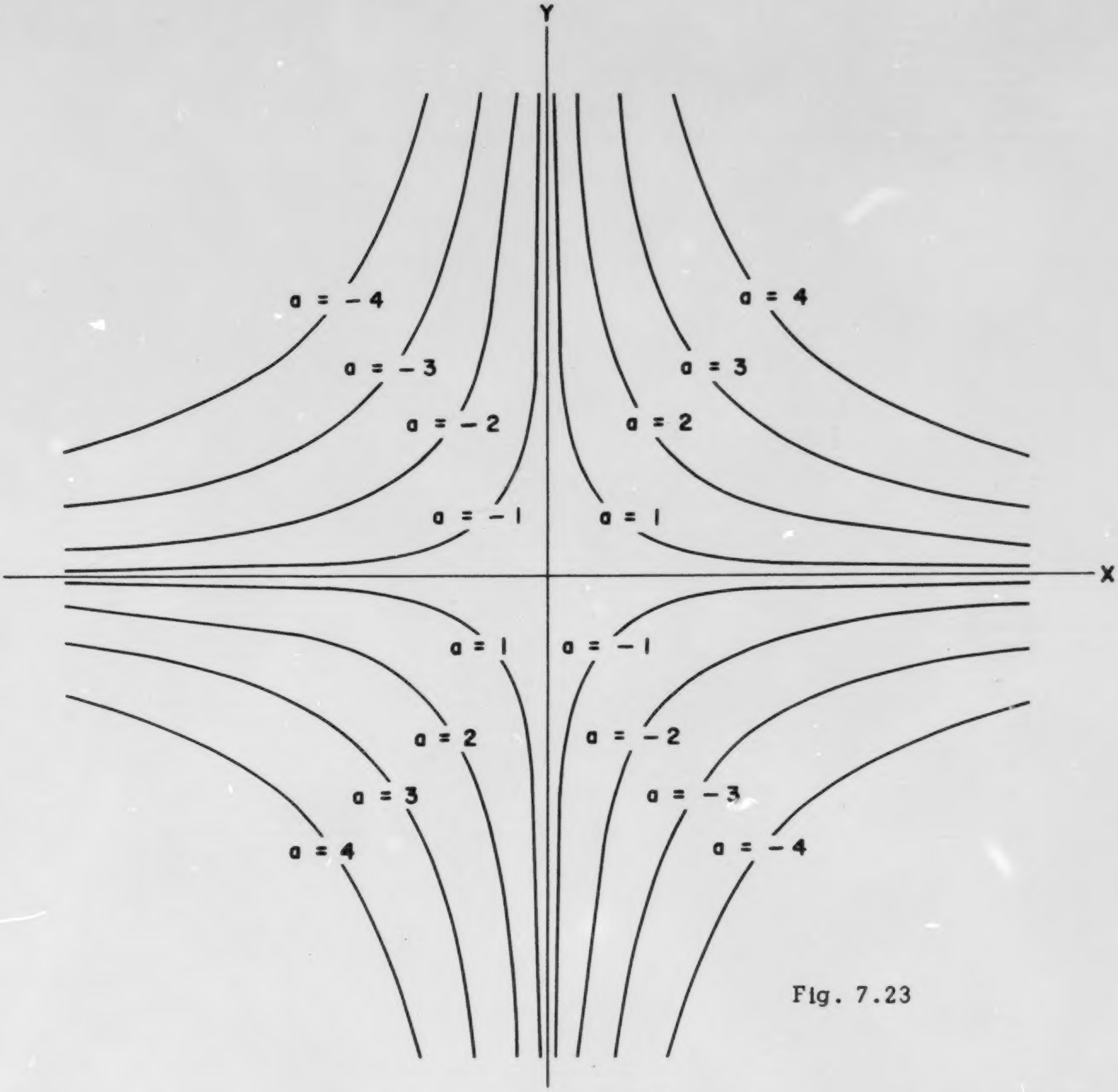


Fig. 7.23

A line which a curve approaches but never meets is called an *asymptote*. As you can see from Fig. 7.22, not only is a curve representing $y = \frac{a}{x}$ asymptotic to the *x* axis; it is also asymptotic to the *y* axis. To see this, consider the equation $y = \frac{a}{x}$, rewritten as $x = \frac{a}{y}$. If *y* is taken as the independent variable, increasing positive values of *y* lead to decreasing values of *x*, but *x* never becomes zero and the curve is asymptotic to the y axis. The same is true for increasing negative values of y. The intersection of the asymptotes of a hyperbola (in this case the origin of the coordinate system) is called the center of the hyperbola.

Since the function $x = \frac{a}{y}$ never equals zero, no matter how large y is made, the equation of a hyperbola in the form $y = \frac{a}{x}$ is undefined for the particular value x = 0. We say that the function has a singularity at the value of x that lies on a vertical asymptote.

Following the procedure developed in Section 7.2, hyperbolas homomorphic to the one described by Equation (9) are given by

$$y - c = f(x - d) = \frac{a}{x - d} \tag{10}$$

All points on the graph displaying this function are displaced d units horizontally and c units vertically relative to the corresponding points on the graph of $y = \frac{a}{x}$ (Fig. 7.24).

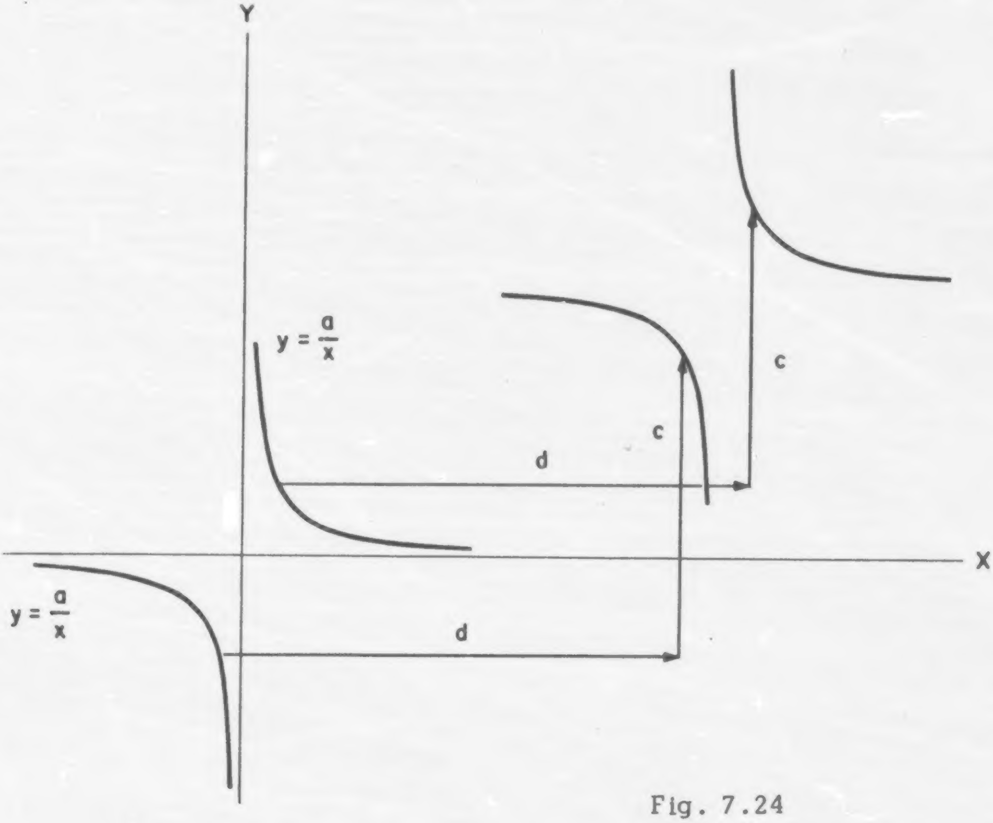


Fig. 7.24

Questions

1. The equation $y - c = \frac{a}{x - d}$ can be rewritten in the standard form

$$yx + mx + ny = g$$

Find the relationship between the parameters a, c, d, and the parameters m, n, g of the two equations.

2. (a) What are the equations for the asymptotes of the hyperbola given by $y - c = \frac{a}{x - d}$ which is homomorphic to the hyperbola described by $y = \frac{a}{x}$?

(b) What are the coordinates of the center of this hyperbola?

(c) Does the value of a in the equation in part (a) affect the asymptotes?

3. (a) Write down the equation for a rectangular hyperbola with the following position of the center C and the following value of the constant a:

(1)	$C = (-2,3)$	$a = 1.5$
(2)	$C = (5,0)$	$a = -3.4$
(3)	$C = (-4,-10)$	$a = -17$

(b) What are the asymptotes for each of the curves in (a)?

4. Show that the graphs of y versus x corresponding to the following relations are rectangular hyperbolas. Specify in each case the center and the asymptotes and then sketch the graphs:

(a) $y + 3 = \frac{4}{x - 2}$

(b) $y = \frac{6}{x - 5} + 4$

(c) $xy = 3$

(d) $xy - x = -2$

5. For which values of x in the function $y = \frac{3}{x}$ will the vertical distance of the graph from the x axis be less than

   (a) 1

   (b) 10^{-25}

   (c) $10^{-1,000,000}$

   Hint: How are coordinates of points on the graph related to vertical distance from axes?

6. (a) For which values of x will the distance from the y axis be less than

   (1) 1

   (2) 10^{-5}

   (3) $10^{-1,000,000}$

   (b) What are the corresponding values of y?

7.9 The Inverse Square Function

When the relation between the independent variable x and the dependent variable y is given by

$$y = \frac{a}{x^2} \tag{11}$$

we say that y is proportional to the inverse square of x. This means that if x is made n times as large as some initial value, y will be $\frac{1}{n^2}$ times the initial value of y.

Like the function $y = \frac{a}{x}$, the function $y = \frac{a}{x^2}$ is defined for all values of x except $x = 0$. Therefore, graphs of $y = \frac{a}{x^2}$ fall in two parts separated by the y axis as an asymptote. Since the same value for y is obtained whether you insert x or -x in $y = \frac{a}{x^2}$, the two parts are symmetric about the y axis (Fig. 7.25).

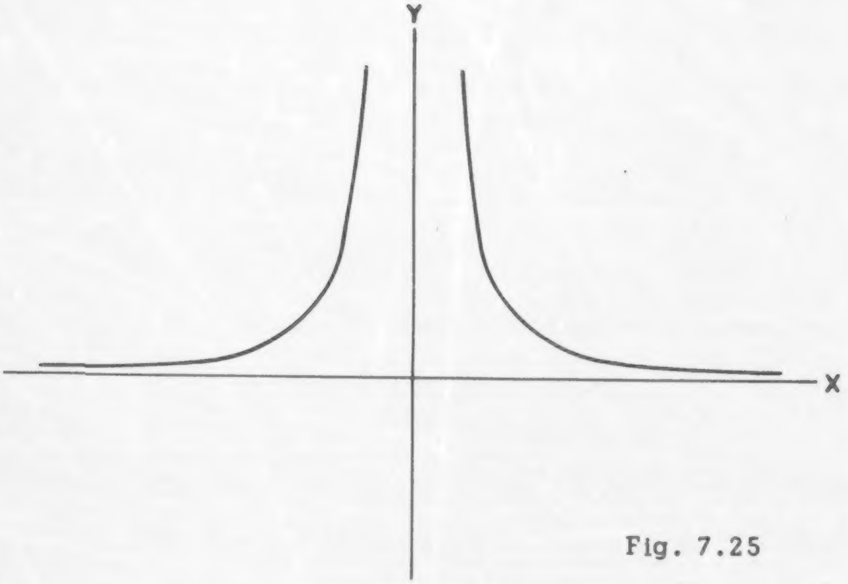


Fig. 7.25

For larger negative or positive values of x the corresponding value of y will get closer to zero, but will never equal zero, no matter how large x becomes. Therefore, the x axis is an asymptote of the curve. Since the value of x is equal to the distance of the curve from the y axis, this distance can be as small as we wish if we just go to sufficiently large (positive or negative) values for y. It means that the y axis is an asymptote of the curve — i.e., $y = \frac{a}{x^2}$ has a singularity at $x = 0$.

Since x^2 is positive for both positive and negative values of x, y will always have the same sign as a. Thus for positive values of a, the curve will lie above the x axis and for negative values of a, it will lie below the x axis. Figure 7.26 is a family of graphs of $y = \frac{a}{x^2}$ corresponding to different values of a.

Comparing Fig. 7.26 and Fig. 7.23, you can see that the graphs of $y = \frac{a}{x}$ and $y = \frac{a}{x^2}$ both have the x axis and the y axis as asymptotes, but the two parts of the graph $y = \frac{a}{x}$ are symmetric with the line $y = -x$ (if $a > 0$), as the axis of symmetry, whereas neither part of the graph of $y = \frac{a}{x^2}$ is symmetric by itself. However, the complete curve is symmetric about the y axis. For a given value of a, $y = \frac{a}{x}$ and $y = \frac{a}{x^2}$ have the same y value for $x = 1$, that is, their graphs both pass through the point $(1,a)$. For $x > 1$, $x^2 > x$, and so $\frac{a}{x^2}$

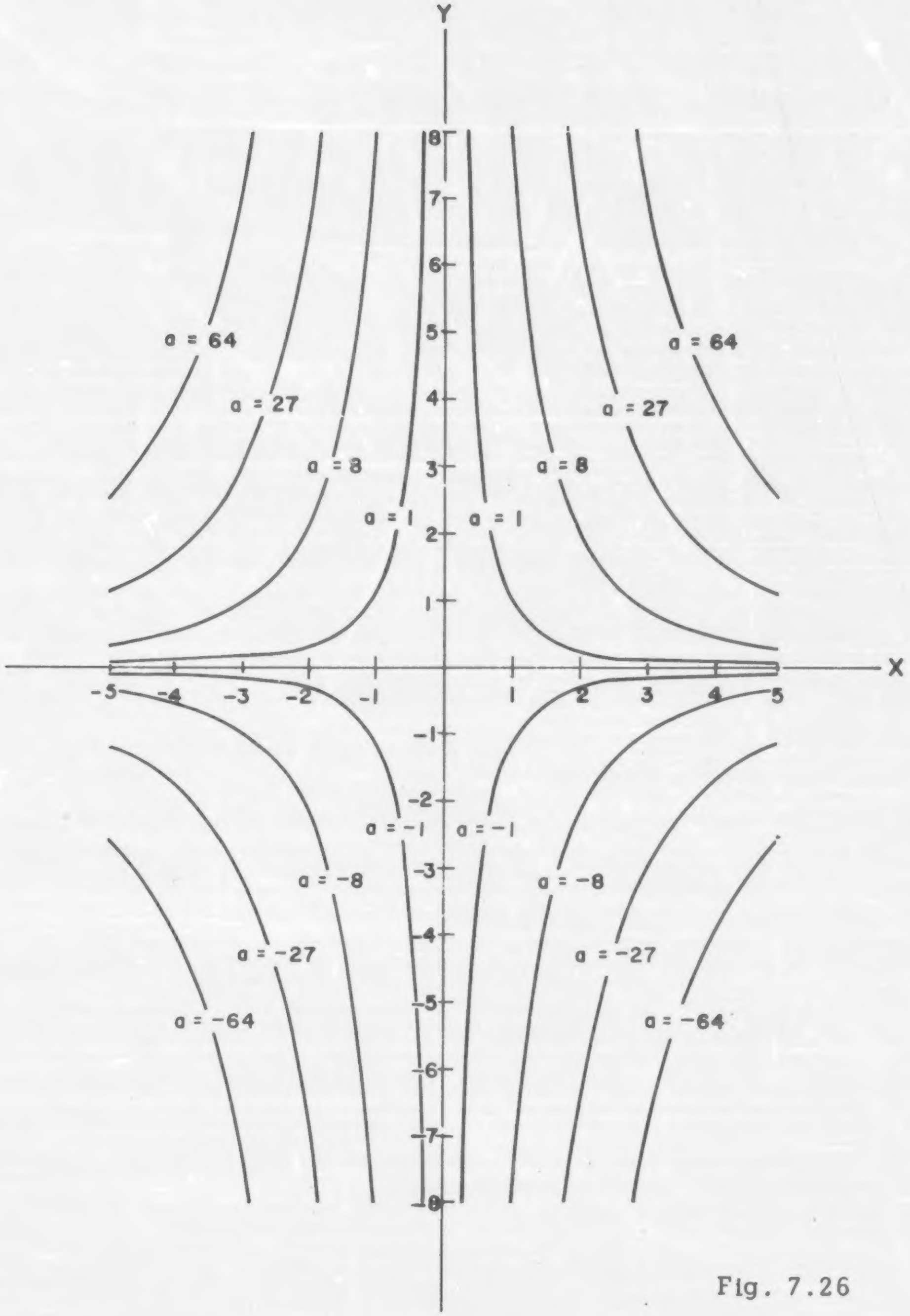


Fig. 7.26

is smaller than $\frac{a}{x}$. But for values of $\underline{x}$ between 0 and 1, $x^2 < x$; hence, in this region, $\frac{a}{x^2}$ is greater than $\frac{a}{x}$, as shown in Fig. 7.27. Notice that the graph of $y = \frac{4}{x}$ approaches the $\underline{x}$ axis and the $\underline{y}$ axis at the same rate, but the graph of $y = \frac{4}{x^2}$ approaches the $\underline{x}$ axis faster than it approaches the $\underline{y}$ axis.

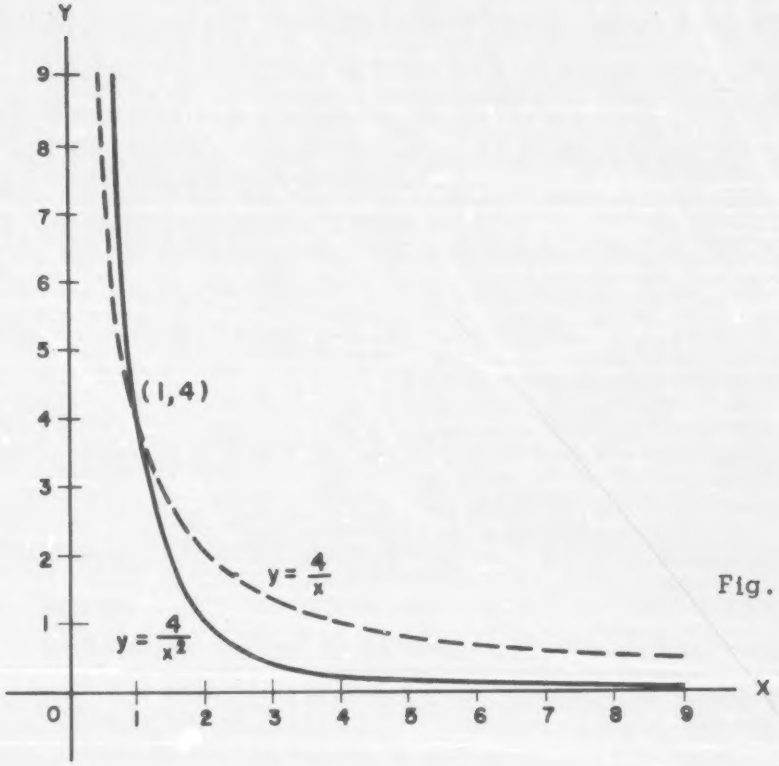


Fig. 7.27

The equation for a graph homomorphic to that of $y = \frac{a}{x^2}$, where the intersection of the asymptotes has the coordinates (d,c) instead of (0,0), is found by replacing x by x - d and y by y - c in the equation $y = \frac{a}{x^2}$, giving

$$y - c = \frac{a}{(x - d)^2}$$

Questions

1. (a) How is the circumference of a circle related to its radius?

(b) Suppose there is a source of particles at the center of the circle and the number of particles crossing a unit of length on the circumference of a circle of radius 1 is n. How many particles will cross a unit length on the circumference of a circle of radius r? (Assume that no particles are lost.)

(c) How is the surface area of a sphere related to its radius?

(d) If particles are emitted evenly in all directions from the center of the sphere, how will the number of particles passing through a unit area of a sphere depend on the radius of the sphere?

2. On the same graph paper sketch very roughly the graphs of the functions $y = \frac{1}{x}$, $y = \frac{1}{x^2}$, $y = \frac{1}{x^3}$, $y = \frac{1}{x^4}$, $y = \frac{1}{x^5}$. Then draw a detailed graph of the same functions for $0 < x < 1$. What do you predict will be the general shape of the graph for $y = \frac{1}{x^{10}}$?

7.10 Recognizing Hyperbolas and Inverse Square Functions

The most characteristic features of graphs corresponding to functions of the type $y = \frac{a}{x}$ and $y = \frac{a}{x^2}$ are the asymptotes. The fact that the graphs have two branches is usually of little practical value since the relationship we are looking for often makes sense only for positive values of the independent variable as in the case of pressure and volume. Therefore, the experimental data will all be on one branch of the graph. If the graph in question seems to have the x axis and the y axis as asymptotes, you can distinguish between the two kinds of functions $y = \frac{a}{x}$ and $y = \frac{a}{x^2}$ by looking for symmetry about the line $y = x$ or $y = -x$, depending on the sign of a (Section 7.8). When considering symmetry it is, of course, important that the same scale be used on the x axis and the y axis.

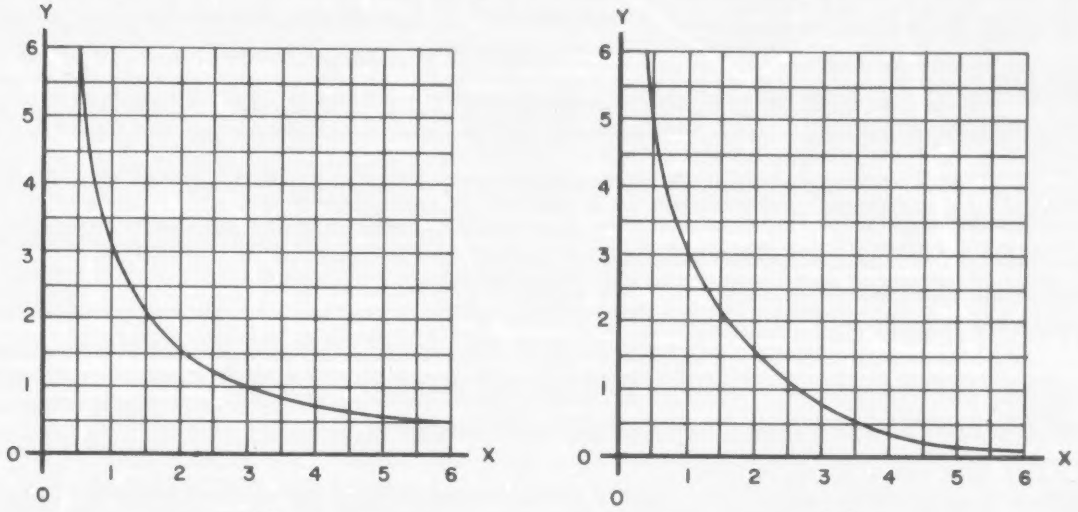
If the graph is symmetrical about one of these lines you have reason to believe that it is probably described by a function of the type $y = \frac{a}{x}$. This can be checked by evaluating the product yx for different points along the curve to see if it remains constant. The value of a is then just the product yx. Or you can calculate $\frac{1}{x}$ for several values of x and graph y as a function of $\frac{1}{x}$ to see if you get a straight line through the origin. If this is the case, you can find the value of a by calculating the slope of the straight line.

If the graph is not symmetrical about the line $y = x$ or the line $y = -x$ you may check for a function of the type $y = \frac{a}{x^2}$, by evaluating the product yx^2 for different points to see if it remains constant. Or you can graph y as a function of $\frac{1}{x^2}$ to see if you get a straight line through the origin. If so, the slope will determine the value of a.

If the x axis and the y axis are not asymptotes to the graph, but some other lines $y = c$ and $x = d$ seem to be asymptotes, you can expect functions of the type $y - c = \frac{a}{x - d}$ or $y - c = \frac{a}{(x - d)^2}$ and test if either type of function describes the curve by plotting $y - c$ as a function of $\frac{1}{x - d}$ or as a function of $\frac{1}{(x - d)^2}$, or by evaluating the product $(y - d)(x - d)$ or $(y - c)(x - d)^2$ for different points on the curve to see if the product remains constant.

Questions

1. For the curves in Fig. 7.28 decide if they are hyperbolas or inverse square functions and if they are either, give their equations.



(a) (b)

Fig. 7.28

2. The general equation for a rectangular hyperbola is $y - c = \frac{a}{(x - d)}$.

   (a) What restriction on the constants is made by requiring the hyperbola to pass through (0,0)?

   (b) If we also require the hyperbola to pass through (1,1) and (2,4), the constants are uniquely determined. Find them.

   (c) Write down the equation of the hyperbola passing through (0,0), (1,1), and (2,4).

   (d) Find the equation of the parabola which also passes through the three points in part (c).

3. Given that a curve has a vertical asymptote at $x = 2$ and passes through the origin, list some simple algebraic equations that it might satisfy.

4. What is a simple algebraic expression which yields a graph asymptotic to the line $y = 2x$ and having a singularity at $x = -1$?

Chapter 8. DERIVATIVES AND ANTIDERIVATIVES

8.1 Function and Slope

In the preceding two chapters, we have seen how a great deal of empirical information can be represented in graphical form, and in some simple cases can be reduced to a mathematical rule in terms of power functions. In this chapter we shall show that, in general, when a function is given in graphical or algebraic form, additional useful information may be extracted from it. To illustrate this point, we shall give the function described by the graph in Fig. 8.1 three different meanings and see what the additional information is in each case.

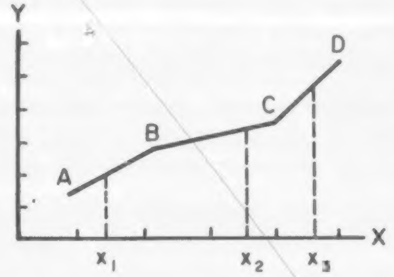


Fig. 8.1

(i) Consider a straight road going up a hill, and let the xy plane be the vertical plane that contains the road. Then the graph in Fig. 8.1 describes the elevation of each point on the road as a function of the horizontal distance. To find the elevation for any particular value of x, we simply read the corresponding value of y off the graph. Thus, for example, the elevation at x_2 is greater than that at x_1. Alternatively, the graph consists of three straight segments, each of which is given by an expression of the form $y = ax + b$. (The values of a and b are different for each segment.) To find the elevation at a given point x_1, we substitute the value of x_1 into the above equation and calculate the corresponding value of y_1.

Suppose now that the question is "How hard is it to push a cart up the road?" The amount of push we have to exert does not depend on the elevation but on the steepness of the road — that is, its slope. From Fig. 8.1 it is evident, therefore, that it is harder to push up a cart at x_1 than at x_2. The slope of the segment AB is greater than the slope of the segment BC.

(ii) Let the horizontal coordinate in Fig. 8.1 represent time, and the vertical coordinate represent position of a car on a road. Then the graph gives the position of the car as a function of time. We can also learn from the graph (or the equivalent expression $y = ax + b$) how fast the car is moving at any moment. Since velocity is the rate of change of position per unit time, it is given by the slope of the position-time graph. (See Section 7.5.) From the graph we see that the car moves faster at time x_3 than at x_2, and moves at some intermediate velocity at x_1. The fact that at x_1 it is farther away from the starting point has no bearing on its velocity; the slope or rate of change of a function contains information different from that of the function itself.

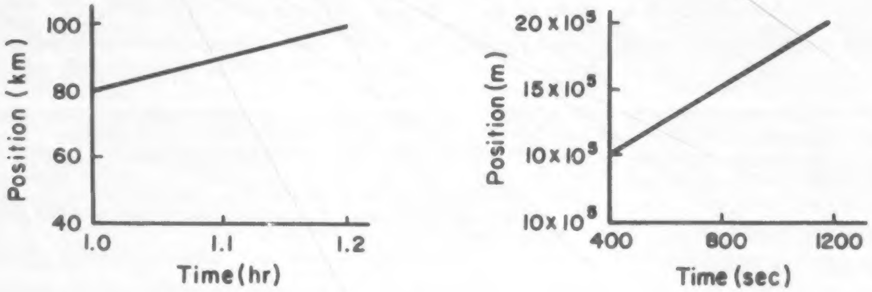
(iii) Now let the horizontal axis in Fig. 8.1 represent time and the vertical axis the cost of living (i.e., the cost of a specified list of goods and services). The graph then gives the cost of living as a function of time. If the cost of living is constant, it is taken for granted. When it goes up, people begin to complain. When it goes up fast, people are likely to complain more. During which period do you think the population was most irritated? Here again we have an example where a quantity of interest is not given by the function itself but by its rate of change, which can be easily extracted from the graph if the graph is made up of straight line segments.

Whatever the graph in Fig. 8.1 is meant to represent, the slope of each straight line segment is given by the coefficient a in $y = ax + b$. As was shown in Section 7.5, $a = \frac{\Delta y}{\Delta x}$.

Questions

1. Give a possible set of units for the independent variable, the dependent variable and the slope of the graph in Fig. 8.1 for each of the three examples cited in the text.

2. The position vs. time graphs of two cars are given in Fig. 8.2(a) and 8.2(b). Which car is moving faster?



(a) Fig. 8.2 (b)

8.2 The Slope Function

Most functions we encounter in applications are not linear functions, that is, they are not represented by straight lines. However, even in the case of curves we have an intuitive feeling for slope: Everybody will agree that the parabola $y = x^2$ in Fig. 8.3 becomes steeper as $\underline{x}$ increases. What we need is a way to translate the intuitive feeling into a clearly defined measure for the slope of a curve at a given point.

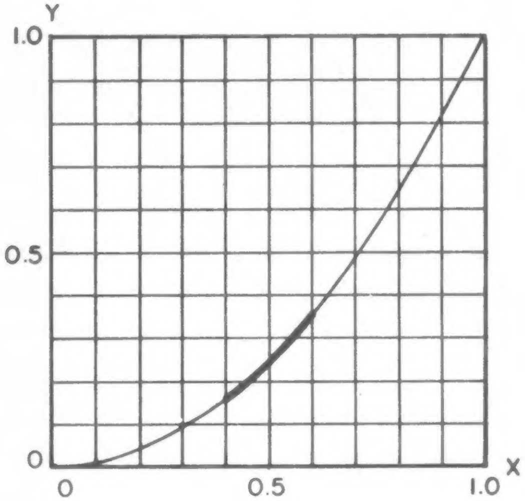
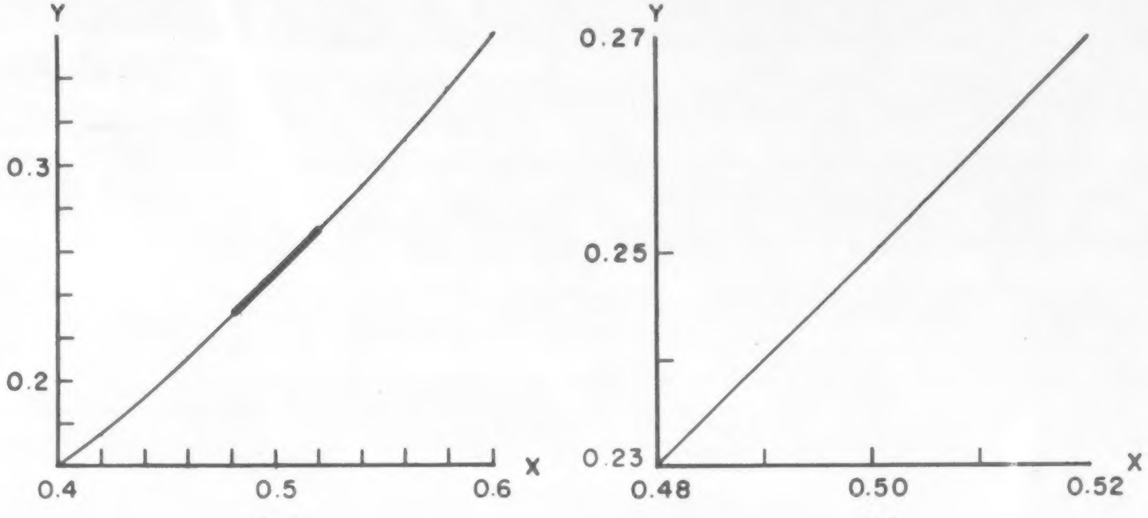


Fig. 8.3

Let us start with the point $x = 0.5$. To characterize the slope of the curve at this point we look at the point and its vicinity with a magnifying glass. This, in effect, is done in Fig. 8.4(a), which shows the heavy portion of Fig. 8.3, for $0.40 \leq x \leq 0.60$, magnified by a factor of 5. Note that this segment of the curve looks much more like a straight line. Repeating the process, an additional magnification by a factor of 10 yields Fig. 8.4(b), which covers only the region corresponding to $0.48 \leq x \leq 0.52$. This segment of the curve now appears to be very nearly a straight line, which upon inspection turns out to have slope $a = \frac{\Delta y}{\Delta x}$ very close to 1.0.



(a) (b)

Fig. 8.4

Therefore, it appears that an answer to the question "What is the slope of the curve $y = x^2$ at $x = 0.5$?" is to say: "It is the same as that of a straight line of slope 1.0."

This magnification process is of course cumbersome, and it would be inconvenient to have to carry it out in practice every time we wanted to know the rate of change of a function $y = f(x)$ at some value of $\underline{x}$. However, we are spared from having to do this by the simple observation that the magnified small segment of the curve has almost the same slope as the straight line which is geometrically tangent to the curve at any point (x_1, y_1). In Fig. 8.5(a), (b), and (c) we have added the tangent lines to the sections of the curve shown in Fig. 8.3 and 8.4(a) and (b). We see that the curve does indeed become less and less distinguishable from the tangent line.

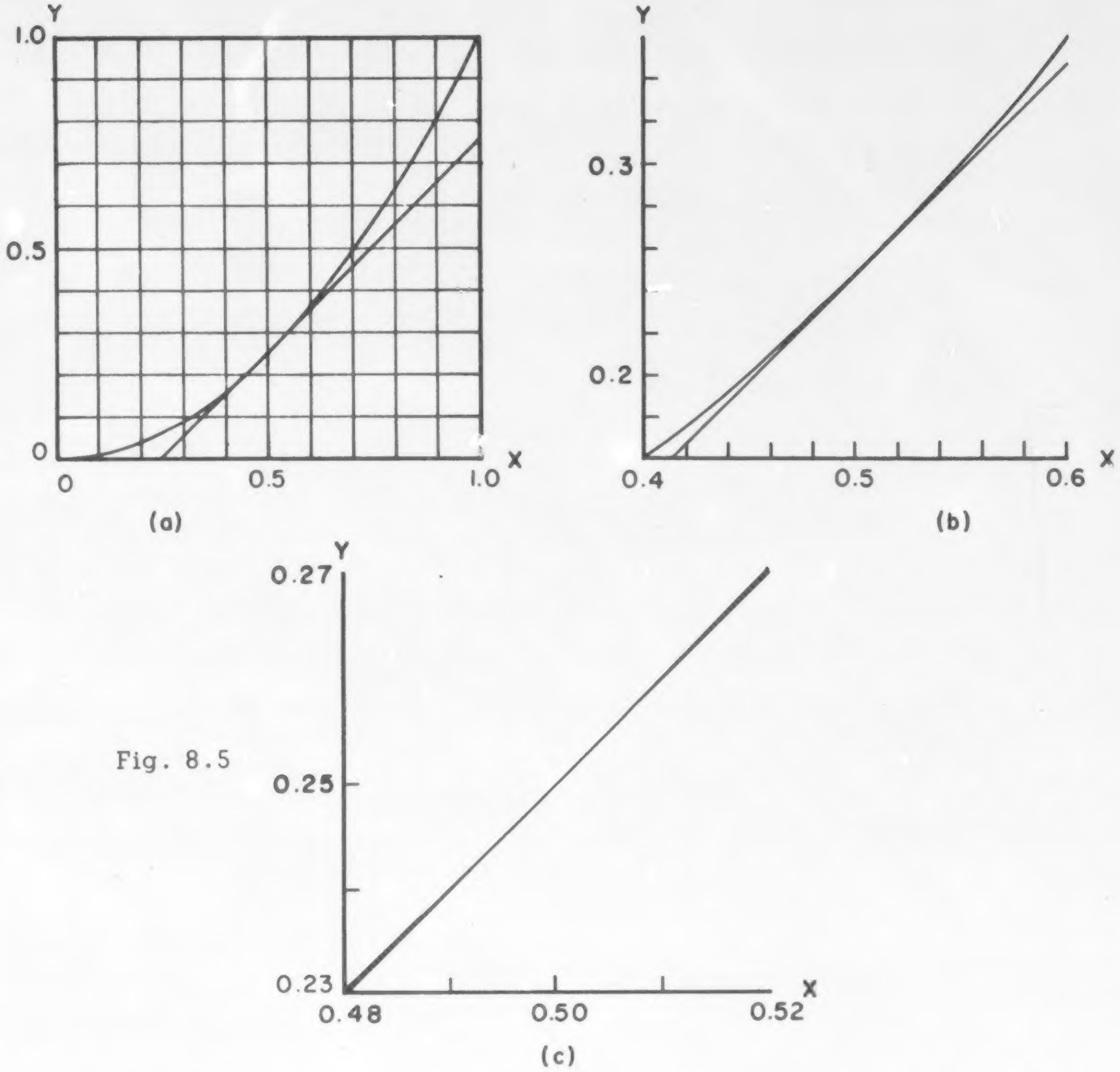


Fig. 8.5

Therefore we define the slope of a smooth curve at a point (x_1, y_1) to be the slope of the tangent line to the curve at that point. This gives the rate at which y is changing with respect to x at $x = x_1$.

If a function $y = f(x)$ is graphed, a practical way of finding its slope at x_1 is to lay a ruler down on the graph and adjust its position until it touches the curve at the point $(x_1, f(x_1))$ at the correct angle, and then draw the tangent line and measure $a = \frac{\Delta y}{\Delta x}$ for this line. Some error in measurement will of course be present, its size depending on how carefully we draw and measure the tangent line, and of course on the accuracy to which the curve itself can be drawn from the given data. In Fig. 8.5(a), for example, we might judge that the points (0.25,0) and (1.00,0.75) lie on the line tangent at (0.5,0.25). This gives a slope value of $a = \frac{\Delta y}{\Delta x} = \frac{0.75}{0.75} = 1.00$.

We can repeat this process for any point on the curve. If for each point on the curve there is only one tangent, then the slope of the tangent is a function of the independent variable x. Accordingly this function can be called the slope function. Frequently the notation $f'(x)$ is used to denote the slope function of $f(x)$. Since the slope function is derived from $f(x)$ it is more often called "the derivative of $f(x)$."

A word of caution is in order on what is meant by the line geometrically tangent to a curve $y = f(x)$ at some abscissa $x = x_1$. Several examples of tangents are shown in Fig. 8.6. The tangent line is sometimes defined as the line which touches the curve but does not cross it at the point in question.

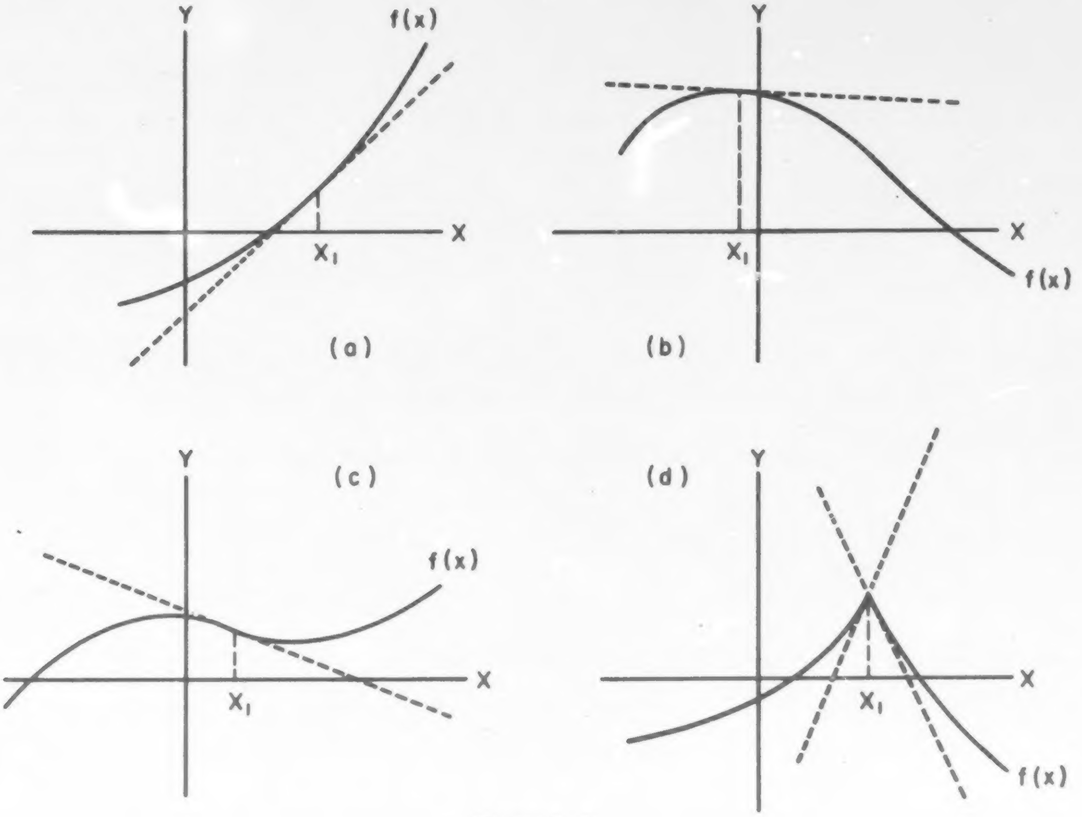


Fig. 8.6

This is correct only in cases like those shown in Fig. 8.6(a) and 8.6(b). However, Fig. 8.6(c) shows that the line tangent to a smooth curve can cross the curve at the point of tangency. Figure 8.6(d) shows the necessity of specifying that the curve be smooth. At $x = x_1$ the curve has a corner, and hence no single tangent exists there.

Questions

1. Figure 6.2 on page 119 is a graph of the population of the United States as a function of time. Use a ruler to draw the tangent to the curve at x = 1800, 1850, and 1900 and find the rate of growth of the population at these times.
2. There are, of course, errors present in the growth rates you measured in Question 1. Investigate the error resulting from constructing and measuring the slope of the tangent line for x = 1850. (Assume that errors made in drawing the curve are negligible.) Draw a line that is just barely too steep, and one that is not quite steep enough, and measure their slopes. From this deduce the uncertainty in your value of the 1840 growth rate.
3. Repeat the work of Question 1 for the year 1850 only, on the alternative version of the population curve, Fig. 6.3, where the horizontal scale has been compressed. Why is it that the value for the slope of the curve at x = 1850 comes out about the same as in Question 1, when the curve appears much steeper at that place?
4. In the case of graphs drawn from data with errors present, special care is needed in measuring slopes. An example is shown in Fig. 6.10 and 6.11 on page 129, where the data were obtained by measuring masses and volumes of various samples of a certain metal.
   (a) Explain why the technique used by the person who drew the graph of Fig. 6.10 is very bad if slope information is needed.
   (b) Does your intuition suggest anything about the slope the curve should have? Would this information be helpful in drawing the curve?

5. By graphically measuring the slope of the function $y = x^2$ (Fig. 8.3), plot the slope function. Can you guess from your result the algebraic formula for the slope function of this curve?

6. By graphically measuring the slope at a number of points, plot the slope function of the function shown in Fig. 8.7. How is the slope function of $f'(x)$ related to the function $f(x)$ itself?

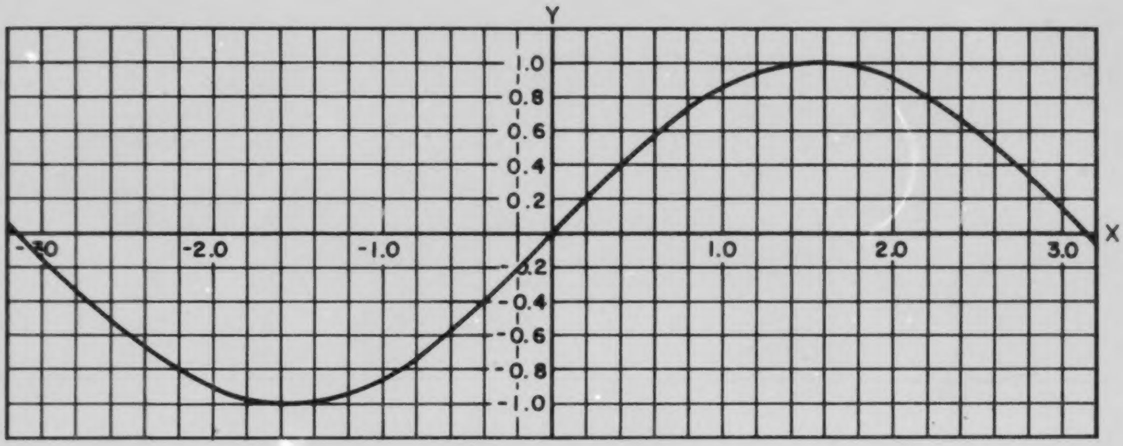


Fig. 8.7

7. Suppose you are steering a boat, trying to keep it on a steady course. There is involved in this situation a function, namely the compass heading as a function of time. In what way do you make use of the value of this function, and in what way do you make use of the value of its slope?

8. (a) Sketch a curve corresponding to a function with the following properties: at $x = x_1$, $f'(x_1) > 0$, for $x > x_1$ the derivative gradually decreases. For $x_2 > x_1$, $f'(x_2) = 0$, and for $x > x_2$ the derivative continues to decrease and is, therefore, negative.

(b) Does the value of $f(x_2)$ have a special significance?

(c) Must any point on a curve for which $f'(x) = 0$ have the same significance?

9. (a) Consider two smooth functions $f(x)$ and $g(x)$, and their derivatives $f'(x)$ and $g'(x)$. If $f'(x) > g'(x)$ over the interval $0 \leq x \leq 10$, does it

follow that $f(x) > g(x)$ over the same interval? Use a sketch to explain your answer.

(b) In country X the rate of inflation was higher than in country Y over the same ten-year period. Was the cost of living also higher in country X than in country Y?

8.3 The Delta Process

So far we have seen how to find values of the slope function by drawing tangents. If we know the algebraic rule (also called the analytical expression) for a function, we can calculate the values of the slope function or derivative without drawing a graph at all. To see this we return to Fig. 8.5(c). There the slope of the tangent at $x = 0.50$ was given by

$$a = \frac{\Delta y}{\Delta x}$$

where $\Delta y = y_2 - y_1$, relates to any two points on the tangent, We can, in particular, choose for y_1 the value corresponding to $x_1 = 0.50$, which is of course the same as the value of the function $y = x^2$ at that point; that is, $y_1 = 0.25$. If we now choose a sufficiently small Δx, say $\Delta x = 0.02$, then the value of y_2 on the tangent corresponding to $x_2 = 0.52$ will be almost the same as the value $y_2 = f(x_2) = {x_2}^2$ on the curve. (See Fig. 8.5(c).) Thus, the slope of the tangent is given approximately by

$$a \approx \frac{f(x_1+\Delta x) - f(x_1)}{\Delta x} \tag{1}$$

The value of the right-hand side of Equation (1) can be calculated directly from the algebraic rule defining $f(x)$; there is no need to draw the graph first.

To improve the approximation we can reduce the size of Δx. As Δx decreases, the value of the y coordinate of the point on the tangent will approach the value of the y coordinate of the corresponding point on the curve as shown in Fig. 8.5. Thus, the error that is introduced by taking the y coordinate of the point on the curve instead of on the tangent will decrease to zero, and the value of the right-hand side of Equation (1) will approach the slope of the tangent. This process is illustrated in Table 8.1, for $x_1 = 0.5$ and various values of Δx.

Table 8.1 may suggest to you that as Δx becomes closer and closer to zero the slope will get closer and closer to 1.000..... . To put this in standard mathematical language, it is likely that as Δx approaches zero, the value of $\frac{f(x_1+\Delta x) - f(x_1)}{\Delta x}$ approaches a limit which gives the value of $f'(x_1)$:

$$f'(x_1) = \lim_{\Delta x \to 0} \frac{f(x_1+\Delta x) - f(x_1)}{\Delta x} \tag{2}$$

The notation $\lim_{\Delta x \to 0}$ is read "the limit as Δx approaches zero of ... "

TABLE 8.1

Δx	$f(x_1)$	$f(x_1+\Delta x)$	$\frac{f(x_1+\Delta x) - f(x_1)}{\Delta x}$
0.1	0.25	0.3600000	1.1000
0.05	0.25	0.3025000	1.0500
0.01	0.25	0.2601000	1.0100
0.005	0.25	0.2550250	1.0050
0.001	0.25	0.2510010	1.0010
-0.1	0.25	0.1600000	0.9000
-0.05	0.25	0.2025000	0.9500
-0.01	0.25	0.2401000	0.9900
-0.005	0.25	0.2450250	0.9950
-0.001	0.25	0.2490010	0.9990

However, from the table alone it is impossible to be sure that the slope will not approach 1.00001, at least not without extending the table until we reach a value closer to 1.00000. What is needed now is a way to calculate this limit without resorting to a long table such as Table 8.1.

Substituting $\Delta x = 0$ directly into Equation (2) will not do, because it will yield the meaningless expression of $\frac{0}{0}$. The way to do it, which is called the delta process, consists of three setps.

The first step is to construct the ratio on the right-hand side of Equation (2) for the specific function whose derivative you wish to find. In our case, where $f(x) = x^2$, this gives

$$\frac{(x_1+\Delta x)^2 - x_1{}^2}{\Delta x}$$

The second step may consist of either of two operations. If possible, the form of the numerator is changed in such a way that a factor Δx can be extracted in order to cancel the Δx in the denominator. After the Δx is canceled, we are sure to avoid the meaningless expression $\frac{0}{0}$ when Δx goes to zero. In this case, a third step is particularly simple: it consists of setting $\Delta x = 0$.*

Let us now carry out the second step for $f(x) = x^2$:

$$(x_1+\Delta x)^2 - x_1{}^2 = x_1{}^2 + 2x_1\Delta x + \Delta x^2 - x_1{}^2 = \Delta x(2x_1+\Delta x)$$

Equation (2) now becomes

$$f'(x_1) = \lim_{\Delta x \to 0} \frac{\Delta x(2x_1+\Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \left(2x_1+\Delta x\right)$$

Now there is no longer an obstacle to setting $\Delta x = 0$, which is the third step of the delta process. In this case we get $f'(x_1) = 2x_1$. For $x_1 = 0.5$, we have $f'(x_1) = 1$ exactly, as was indeed suggested by Table 8.1.

The delta process works for any value of the variable. Thus we shall omit the subscript and write in general that for $f(x) = x^2$ the derivative, $f'(x) = 2x$, or, in short,

$$[x^2]' = 2x \tag{3}$$

where [ ]' stands for the derivative of the function in the brackets.

Questions

1. Using the delta process, find the derivative of $f(x) = 5x^2 - 6x$.
2. The derivative of $f(x) = x^3$ is defined as

$$f'(x) = \lim_{\Delta x \to 0} \frac{(x+\Delta x)^3 - x^3}{\Delta x}$$

   (a) Change the numerator in such a way as to have a factor Δx in it.
   (b) After canceling the Δx, what is the derivative of x^3?

---

*If a factor of Δx cannot be extracted from the numerator, then the form of the numerator must somehow be changed in such a way as to make it possible to find the limit of the ratio as $\Delta x \to 0$. In this chapter we shall look only at cases where a factor Δx can be extracted. In the next two chapters we shall see examples where the limit will be found even though a factor Δx cannot be extracted.

3. Generalize the result of Question 2 to find the derivative of $f(x) = x^n$ where n is a positive integer. (Hint: When you multiply out $(x + \Delta x)^n$, you will get a term x^n and terms with powers of Δx ranging from Δx to Δx^n. Why do you have to know only the coefficient of Δx?)

4. Consider the function

$$f(x) = \begin{cases} x^2 \text{ for } x \leq 1 \\ 3x \text{ for } x > 1 \end{cases}$$

(a) Sketch the function for $0 \leq x \leq 2$.

(b) Does this function have a derivative at $x = 1$?

(c) What happens if you try to apply the delta process to find the derivative at $x = 1$?

8.4 The Derivative of $\frac{1}{x}$ and $\sqrt{x}$

The first step of the delta process can be applied to any function. However, the second step may not always be as straightforward as in the examples treated in the preceding section. We shall illustrate this by finding the derivatives of two common functions: $\frac{1}{x}$ and $\sqrt{x}$.

As with any calculation, it is worth while to get some idea as to what to expect. Figure 8.8 is the graph of $y = \frac{1}{x}$. What can we tell about the derivative of $\frac{1}{x}$ from the graph? From the few tangent lines drawn in the figure, it is evident that all tangents slope downward to the right. Hence the derivative of $\frac{1}{x}$ is negative everywhere (except at $x = 0$, which is not in the domain of $\frac{1}{x}$). For large positive and large negative values of x, the tangents tend to become horizontal, i.e., $\left[\frac{1}{x}\right]'$ approaches zero. For x near zero, the tangent points down almost vertically, hence the derivative is very large and negative.

Now we apply the delta process to find the function that has the features we just described. For $f(x) = \frac{1}{x}$ the first step yields

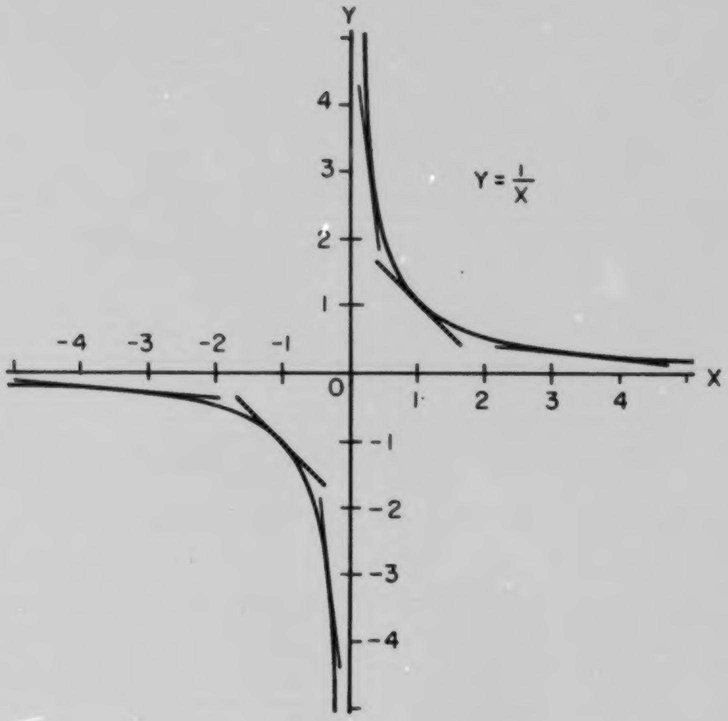


Fig. 8.8

$$\frac{1}{\Delta x}\left[\frac{1}{x + \Delta x} - \frac{1}{x}\right]$$

(For ease of writing, we have put the $\frac{1}{\Delta x}$ in front of the bracket instead of Δx underneath.) Carrying out the subtraction, using a common denominator, gives

$$\frac{1}{\Delta x} \cdot \frac{x - (x + \Delta x)}{(x + \Delta x)x} = \frac{1}{\Delta x} \cdot \frac{-\Delta x}{(x + \Delta x)x} = -\frac{1}{(x + \Delta x)x}$$

Now the third step can be taken by letting $\Delta x = 0$, which yields

$$\left[\frac{1}{x}\right]' = -\frac{1}{x^2} \tag{4}$$

The function $-\frac{1}{x^2}$ indeed has the properties which we predicted from the graph in Fig. 8.8.

What can we predict about the derivative of $\sqrt{x}$ from the graph of this function shown in Fig. 8.9? The derivative is positive, but steadily decreasing as x increases.

To find the derivative of $\sqrt{x}$, we start with

$$\frac{\sqrt{x + \Delta x} - \sqrt{x}}{\Delta x}$$

To be able to get a factor Δx in the numerator requires replacing $\sqrt{x + \Delta x} - \sqrt{x}$ by $(x + \Delta x) - x = \Delta x$ without changing the values of the quotient. This is accomplished by multiplying the numerator and the denominator by $\sqrt{x + \Delta x} + \sqrt{x}$ (see Appendix, page):

$$\frac{\sqrt{x + \Delta x} - \sqrt{x}}{\Delta x} = \frac{\sqrt{x + \Delta x} + \sqrt{x}}{\sqrt{x + \Delta x} + \sqrt{x}} \cdot \frac{\sqrt{x + \Delta x} - \sqrt{x}}{\Delta x}$$

$$= \frac{(x + \Delta x) - x}{(\sqrt{x + \Delta x} + \sqrt{x})\ \Delta x} = \frac{1}{\sqrt{x + \Delta x} + \sqrt{x}}$$

This completes the second step of the delta process. The third step is straightforward: for $\Delta x = 0$ we have

$$[\sqrt{x}]' = \frac{1}{2\sqrt{x}} \qquad (5)$$

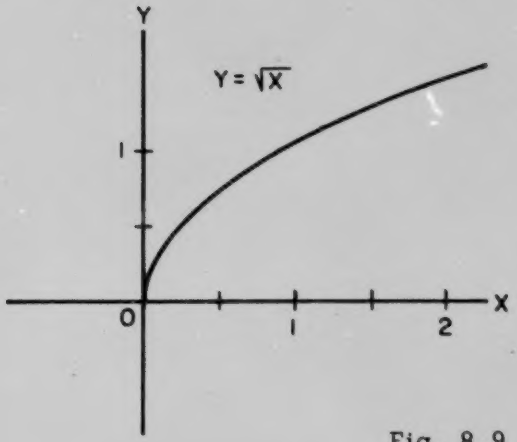


Fig. 8.9

Questions

1. Although the derivative of $\sqrt{x}$ is not defined at $x = 0$, what can you say about the direction of the tangent to the graph of $y = \sqrt{x}$ at $x = 0$?

2. Suppose the graph of $y = \frac{1}{x}$ is drawn using the same scales on the x and y axes. For which values of x will the tangent of the curve make an angle of 135° with the positive direction of the x axis?

3. Noting that $\frac{1}{x} = x^{-1}$ and $\sqrt{x} = x^{1/2}$, can you suggest (without proof) a further generalization for the derivative of $f(x) = x^n$ for negative as well as non-integer values of n? (See Question 3 in the preceding section.)

8.5 Some Properties of Derivatives

If we had to apply the delta process whenever we wished to find the derivative of a given function, it would indeed be very tedious. Fortunately, this is not necessary, as the following two theorems will show.

(i) If the $f'(x)$ is the derivative of $f(x)$, then the derivative of $h(x) = cf(x)$ is $h'(x) = cf'(x)$, where c is a constant. In words, the derivative of a constant times a function is the constant times the derivative of that function. For example, $[5x^3]' = 5 \cdot 3x^2 = 15x^2$.

To prove this theorem we proceed as follows:

$$h'(x) = \lim_{\Delta x \to 0} \frac{h(x + \Delta x) - h(x)}{\Delta x}$$

In general, if $h(x) = cf(x)$, then

$$h'(x) = \lim_{\Delta x \to 0} \frac{cf(x + \Delta x) - cf(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} c\left(\frac{f(x + \Delta x) - f(x)}{\Delta x}\right)$$

The factor c is a constant and does not depend on Δx. We can, therefore, apply the delta process to the ratio in the parentheses and then set $\Delta x = 0$, which yields $f'(x)$. Hence

$$h'(x) = [cf(x)]' = cf'(x) \tag{6}$$

(ii) If $h(x) = f(x) + g(x)$, then $h'(x) = f'(x) + g'(x)$. In words: the derivative of a sum of functions is the sum of the derivatives of the individual functions. For example: $[x^2 + \sqrt{x}]' = 2x + \frac{1}{2\sqrt{x}}$.

The proof proceeds along lines similar to those of the proof of the preceding theorem:

$$[f(x) + g(x)]' = \lim_{\Delta x \to 0} \left(\frac{f(x+\Delta x) + g(x+\Delta x) - f(x) - g(x)}{\Delta x} \right)$$

By rearranging the two middle terms on the right-hand side, we have

$$[f(x) + g(x)]' = \lim_{\Delta x \to 0} \left(\frac{f(x+\Delta x) - f(x)}{\Delta x} + \frac{g(x+\Delta x) - g(x)}{\Delta x} \right)$$

We can now carry out the delta process separately for each ratio, which yields the sum of the derivatives:

$$[f(x) + g(x)]' = f'(x) + g'(x) \qquad (7)$$

Questions

1. (a) Express the area of a circle as a function of its radius.

   (b) What is the rate of change of the area as a function of the radius?

2. What is the derivative of x, $\frac{1}{2}x^2$, $\frac{1}{3}x^3$, $\frac{1}{n}x^n$ $(n \neq 0)$?

3. (a) What is the derivative of $g(x) = c$, where $\underline{c}$ is a constant? Give a geometric reason for your answer. Show that the delta process gives the same result.

   (b) Using the second theorem in this section, what is the derivative of $h(x) = f(x) + c$?

   (c) What can you say about two functions $h(x)$ and $f(x)$, if $h'(x) = f'(x)$ in a given interval?

4. Using the two theorems in this section, prove that $(af(x) + bg(x))' = af'(x) + b'g(x)$, where $\underline{a}$ and $\underline{b}$ are constants. (Hint: Apply the second theorem to the sum, and then apply the first theorem to each term.)

5. Consider the function $f(x) = x + \frac{1}{x}$ in the interval $-5 \leq x \leq 5$ $(x \neq 0)$.

   (a) How is $f(x_1)$ related to $f(x_2)$ where $x_2 = \frac{1}{x_1}$?

   (b) Sketch the graph of $y = f(x)$ in the above interval. Is your answer to part (a) useful in this task?

   (c) At which point is the tangent to the curve horizontal?

   (d) Use the derivative of $f(x)$ to check on your answer to part (c).

8.6 Antiderivatives

Suppose the function $f(x) = 3x^2$ describes the rate at which water flows into a container. What function will describe the volume of water in the container at different times? Or suppose that $f(x) = 3x^2$ describes the slope of a curve as a function of the horizontal coordinate. What expression will describe the curve itself?

The two questions which we have raised are examples of situations where a function $f(x)$ is known and we are looking for another function $F(x)$ whose derivative is $f(x)$, i.e., $f(x) = F'(x)$. The function $F(x)$ is called an antiderivative or integral of $f(x)$. From the preceding sections we know antiderivatives of a number of functions. For example, an antiderivative of $f(x) = 3x^2$ is $F(x) = x^3$. Similarly, an antiderivative of $f(x) = -\frac{1}{x^2}$ is $F(x) = \frac{1}{x}$.

The two theorems about derivatives discussed in Section 8.5 can be restated in terms of antiderivatives and then used to find antiderivatives of related functions.

(i) If $F(x)$ is an antiderivative of $f(x)$ then $cF(x)$ is an antiderivative of $cf(x)$. For example, x^3 is an antiderivative of $3x^2$. Hence $\frac{1}{3}x^3$ is an antiderivative of $\frac{1}{3} \cdot 3x^2 = x^2$. To show the general validity of this theorem we note that for $F'(x) = f(x)$

$$[cF(x)]' = cF'(x) = cf(x)$$

Hence $cF(x)$ is an antiderivative of $cf(x)$.

(ii) An antiderivative of a sum of functions is the sum of the antiderivatives. If $F(x)$ is an antiderivative of $f(x)$ and $G(x)$ is an antiderivative of $g(x)$ then $[F(x) + G(x)]' = F'(x) + G'(x) = f(x) + g(x)$ which proves the theorem.

Questions

1. Find antiderivatives of the functions listed in the table below.

$f(x)$	1	x	x^2	$\sqrt{x}$	$\frac{1}{x^2}$
$F(x)$					

2. Find antiderivatives of the following functions:

   (a) $5x^2$ (c) $\frac{3}{x^2} - x$

   (b) $-\frac{1}{2}\sqrt{x}$ (d) $2\sqrt{x} - 1$

3. Find an integral (antiderivative) of the following function

$$f(x) = k_0 + k_1x + k_2x^2$$

   where k_0, k_1, and k_2 are constants.

4. Find an integral of $f(x) = x^m$. (Hint: first find the derivative of x^{m+1}, then use theorem (i) of this section.)

8.7 The Constant of Integration: The Initial Condition

We have been careful, so far, to speak of an antiderivative of a function f(x) and not the antiderivative. The reason for this is that a function has a whole family of antiderivatives. Specifically, if F(x) is an antiderivative of f(x) so is F(x) + C, where C is any constant. This follows from the second theorem on derivatives: the derivative of a sum is the sum of the derivatives, and from the fact that the derivative of a constant is zero. Thus the antiderivatives of a given function form a family of homomorphic functions displaced vertically with respect to one another. For example: $F(x) = x^3$ is an antiderivative of $f(x) = 3x^2$, since $[x^3]' = 3x^2$ (Fig. 8.10). But so is $x^3 + C$ for any value of C, since $[x^3 + C]' = 3x^2$. The constant C is called the constant of integration.

We now recognize that the question we raised at the beginning of Section 8.6 has no unique answer. The whole family of curves $y = x^3 + C$ has the slope function $f(x) = 3x^2$. Any of them could describe the volume of water as a function of time.

To make the antiderivative unique we have to know its value at some point $x = x_1$. In the case of the water flowing into the container we may know that at $x = 0$ there was no water in the container, i.e., at $x = 0$, $y = 0$. Only the curve $y = x^3$ fulfills this condition and, therefore, uniquely describes the volume of water as a function of time. Similarly, if $x = 1$, $y = 2$, then only the curve $y = x^3 + 1$ satisfies this condition (Fig. 8.10). In gen-

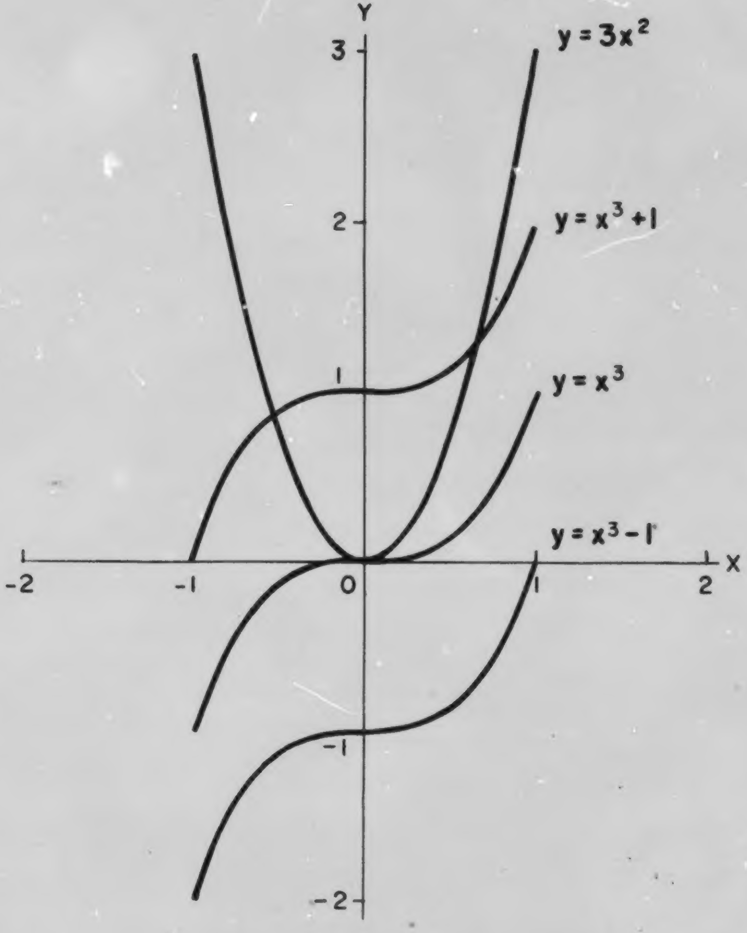


Fig. 8.10

eral, if we specify the value of an antiderivative for a given value of the independent variable, we thereby select a unique member of the family of homomorphic functions. This condition, known as the initial condition, determines the value of the constant of integration. Let F(x) be any antiderivative of f(x). Then the condition $y_1 = F(x_1) + C$, that is $y = y_1$ for $x = x_1$, yields the equation which determines the value of C.

In the case of $f(x) = 3x^2$ we have $F(x) = x^3$, and $y = x^3 + C$ for the family of antiderivatives. If we look for the antiderivative that fulfills the initial condition that for x = 2, y = 10 we have the equation

$$10 = 2^3 + C \qquad \text{or} \qquad C = 2$$

Questions

1. (a) Find the family of antiderivatives of $f(x) = 6x^2 + 2$.

   (b) If you choose the constant of integration to be zero, what is the value of $F(x)$ at $x = 0$, i.e., $F(0)$?

   (c) What must be the value of C if you require $F(0) = 5$?

2. (a) Find an integral $F(x)$, of $f(x) = 5x - \frac{2}{x^2} - 3\sqrt{x}$.

   (b) What is the value of $F(x)$ at $x = 1$, i.e., $F(1)$?

   (c) Give another integral of $f(x)$ for which $F(1) = 2$.

3. Two containers are being filled with water at the same rate over the same time interval. Do they necessarily have the same amount of water in them at the end of that time interval? Relate your answer to Fig. 8.10.

4. The rate of growth of the population of two cities over a period of three years has been the same. Must the two cities have the same population at the end of the third year?

5. What function has the following properties: $F'(x) = 3x$ and $F(1) = -\frac{1}{2}$.

6. An astronaut on the moon throws a rock vertically upward. Suppose the rate at which the rock rises (after leaving the astronaut's hand) is given $v = 10 - 1.5t$ where v is given in meters and t in seconds. What will be the elevation of the rock as a function of time if at $t = 0$ the elevation was $h = 2.0$ meters above the ground? (Note: quite often a function and its derivative are denoted by different letters. In this case $v(t) = h'(t)$. The "v" stands for velocity and "h" stands for height.)

8.8 Short-Range Predictions

Consider a smooth function $f(x)$ and its antiderivative $F(x)$, about which we have the following information:

(a) $f(x_1) = F'(x_1)$ is known, and

(b) $F(x_1)$ is known.

What can you infer about the values of $F(x)$ in the vicinity of x_1? In other words, if we know the value of a function and its antiderivative at a given point, what can we say about the values of the antiderivative nearby?

Figure 8.5 has shown that near a given point the curve corresponding to a smooth function is very close to the line tangent to the curve at that point. The tangent line passing through the point $(x_1, F(x_1))$ has a slope $a = F'(x_1) = f(x_1)$. Hence the linear function corresponding to the tangent line is (see Chapter 7, Equation (7)):

$$\ell(x) = F(x_1) + f(x_1)\,(x - x_1)$$

or, substituting Δx for $x - x_1$, this becomes

$$\ell(x_1 + \Delta x) = F(x_1) + f(x_1)\Delta x$$

For sufficiently small Δx, we have

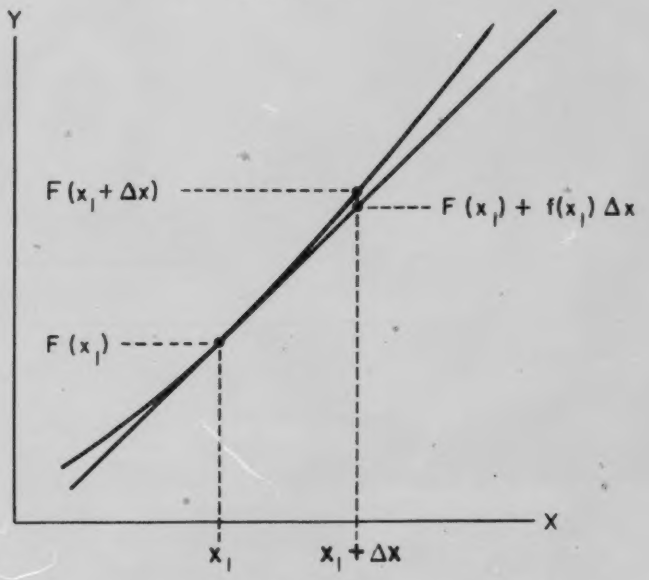
$$F(x_1 + \Delta x) \approx \ell(x_1 + \Delta x)$$

Hence, for any smooth function

$$F(x_1 + \Delta x) \approx F(x_1) + f(x_1)\Delta x \tag{8}$$

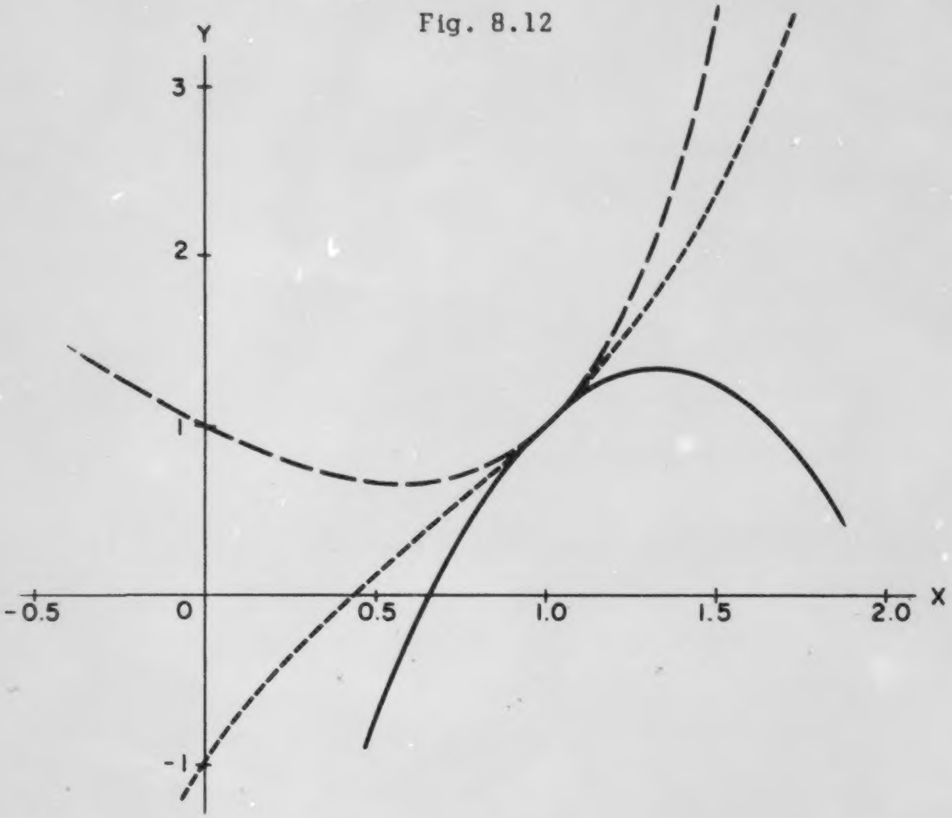
independently of how the function behaves farther away from x_1. (See Fig. 8.11).

Fig. 8.11



How small "sufficiently small" is, will vary from case to case, but for any smooth function for which its value and the value of its derivative are known at one point, we can predict the value of the function near that point. This is illustrated for three different functions in Fig. 8.12. All

Fig. 8.12



three curves pass through the point (1,1) and have a slope of 2 at that point (note that the axes have different scales). But otherwise they correspond to completely different functions. Nevertheless, near x = 1 they are very close to their common tangent line, and Equation (8) can be used to predict the values of any of these functions near x = 1.

This result has many practical applications. For example, suppose an airplane is sighted. Its position and velocity (the rate of change of position as a function of time) are determined at a given moment by radar. One can use Equation (8) to predict where that airplane is going to be a short time later. This is so because the airplane's position is a smooth function of time, even when it changes speed, altitude, or direction. The trajectory of an airplane cannot have a gap or a sharp corner.

Questions

1. At a given instant an airplane is exactly overhead and moving at 250 m/sec due north.

   (a) In relation to Equation (8), what corresponds to $f(x_1)$, $F(x_1)$?

   (b) Where will the airplane be 1 sec later? 10 sec later? Are you equally sure of both answers?

2. A marker on Highway 20 is at an elevation of 850 meters. At that point the highway has a slope of 0.08, rising toward the east. What will be its elevation 150 meters to the east?

8.9 A General Way of Calculating Integrals

In the preceding section we saw that if we know the value of a function and its derivative at x_1, we can calculate the approximate value of the function at $x_1 + \Delta x$ (Equation (8)). Suppose the derivative is known throughout its domain and the function itself is known at x_1. Then we can find an approximate value of the function itself for some other value of x, say x_2, not necessarily near x_1. This is done by a succession of steps, similar to the one expressed by Equation (8). Specifically from the approximate value of $F(x_1 + 2\Delta x)$.

$$F(x_1 + 2\Delta x) \approx F(x_1 + \Delta x) + f(x_1 + \Delta x)\Delta x$$

Substituting from Equation (8) for $F(x_1 + \Delta x)$ yields

$$F(x_1 + 2\Delta x) \approx F(x_1) + f(x_1)\Delta x + f(x_1 + \Delta x)\Delta x$$

We can continue the process and calculate the approximate value of the function itself at $x_1 + 3\Delta x$:

$$\begin{aligned} F(x_1 + 3\Delta x) &\approx F(x_1 + 2\Delta x) + f(x_1 + 2\Delta x)\Delta x \\ &\approx F(x_1) + f(x_1)\Delta x + f(x_1 + \Delta x)\Delta x + f(x_1 + 2\Delta x)\Delta x \\ &= F(x_1) + \sum_{n=0}^{2} f(x_1 + n\Delta x)\Delta x \end{aligned}$$

If we let $x_2 = x_1 + N\Delta x$, we can generalize the process further to find

$$F(x_2) \approx F(x_1) + \sum_{n=0}^{N-1} f(x_1 + n\Delta x)\Delta x \tag{9}$$

This process is illustrated in Fig. 8.13 for the function whose derivative is $f(x) = x$, and the initial condition is that for $x_1 = 0.4$, $F(0.4) = 0.08$. For this simple case, the exact answer is $F(x) = \frac{x^2}{2}$ (Section 8.6). The interval between $x_1 = 0.4$ and $x_2 = 1.6$ is first divided into three parts, i.e., $N = 3$ and $\Delta x = \frac{x_2 - x_1}{N} = 0.4$. As can be seen from Fig. 8.13, the approximation becomes poorer as x increases from x_1 to x_2. It can be improved by decreasing the size of Δx, i.e., increasing the number of steps into which the interval $x_2 - x_1$ is divided. In Fig. 8.14 the same interval has been divided into $N = 6$ steps with $\Delta x = 0.2$. The approximation, as you can see, is better. If we set $\Delta x = 0.1$, dividing the interval into 12 segments, the approximation is even better (Fig. 8.15).

Fig. 8.13

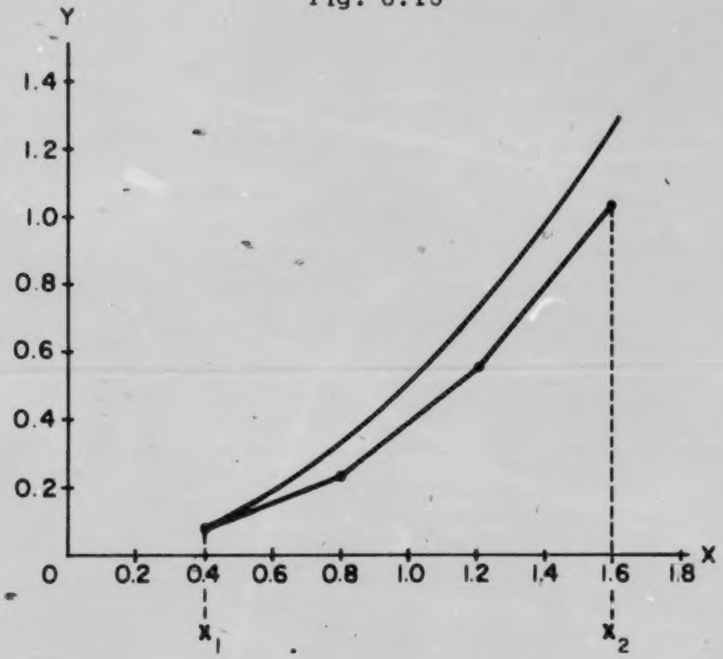


Fig. 8.14

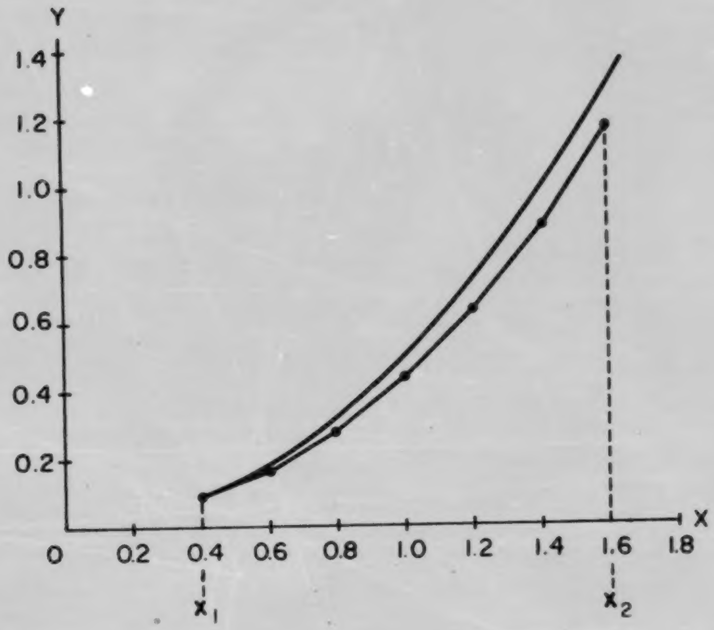
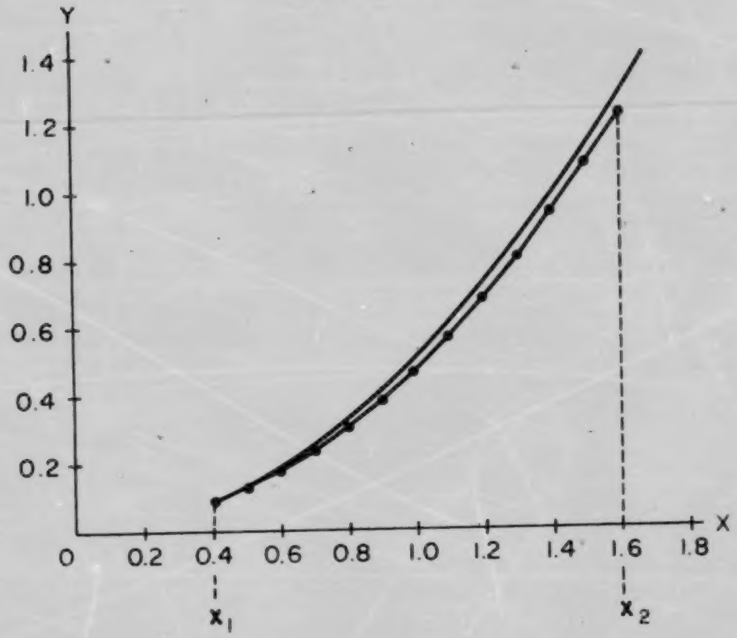


Fig. 8.15



By decreasing the size of Δx, we approximate the curve corresponding to $F(x)$ by an increasingly large number of straight-line segments which become shorter and closer to the curve itself. In the limit as $\Delta x \to 0$ (or $N \to \infty$) we get the curve itself. The endpoint of the last segment is then exactly $F(x_2)$:

$$F(x_2) = F(x_1) + \lim_{\Delta x \to 0} \sum_{n=0}^{N-1} f(x_1 + n\Delta x)\Delta x \qquad (10)$$

There is a generally accepted shorthand notation for the limit in Equation (9): the symbol $\sum$ is changed to a stretched "S"; since in the limit $x_1 + n\Delta x$ takes up all values of x between x_1 and x_2, it is simply replaced by x; finally Δx is replaced by dx. Thus, in shorthand

$$F(x_2) = F(x_1) + \int_{x_1}^{x_2} f(x)dx \qquad (11)$$

which is read as "integral of $f(x)dx$ from x_1 to x_2."

If we can calculate the integral in Equation (10) for any values of x_2, then we have solved the problem of finding a function $F(x)$ whose value is known for one value of x (e.g., x_1) and whose derivative $f(x)$ is known in its entire domain. The limit of the sum given by Equation (11) gives the integral or antiderivative of $f(x)$ for which $F(x_1)$ is specified. The initial condition is built into this method of finding the antiderivative.

Questions

1. The derivative $f(x)$ of a function $F(x)$ is known for the values of the independent variable given in the table below.

x	f(x)	x	f(x)
0	150	2.5	88
0.5	134	3.0	79
1.0	120	3.5	72
1.5	107	4.0	65
2.0	97		

(a) Calculate an approximate value for $F(4)$ using $\Delta x = 1.0$, subject to the initial condition $F(0) = 0$.

(b) Improve the approximation by using $\Delta x = 0.5$.

(c) How would you compare the two approximations?

2. (a) Find an approximate value for $F(1)$, the antiderivative of $f(x) = \frac{1}{1 + x^2}$ subject to the initial condition $F(0) = 0$, by dividing the interval between $x_1 = 0$ and $x_2 = 1$ into five parts.

(b) Repeat part (a) dividing the interval into ten parts.

(c) How do your answers compare?

8.10 The Area Under a Curve

The sum $\sum_{n=0}^{N-1} f(x_1 + n\Delta x)\Delta x$ in Equation (9) has a simple geometrical interpretation. Each term in the sum equals the area of a rectangle whose base is Δx and whose height is $f(x_1 + n\Delta x)$. If we draw the graph of $y = f(x)$, then these rectangles touch the graph at their left corner. This is illustrated in Fig. 8.16 for $N = 4$. Here the sum of the areas of the rectangles is a rough approximation for the area under the curve.

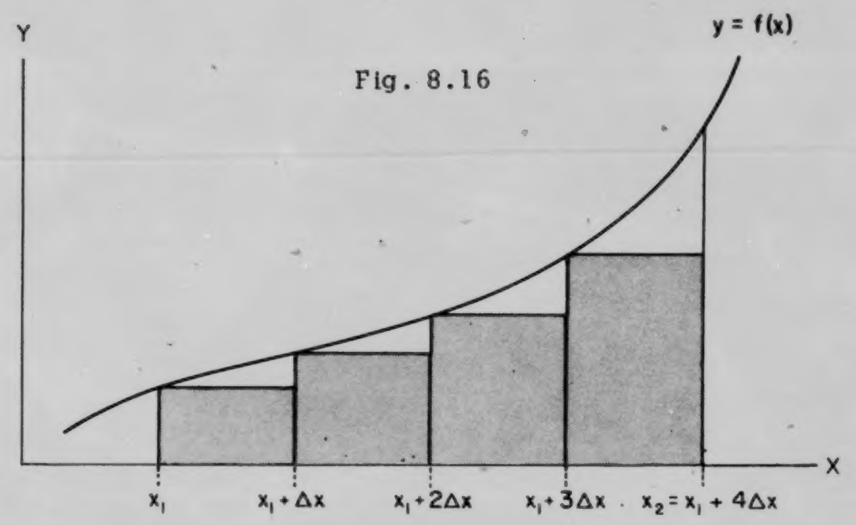


Fig. 8.16

If we divide the interval $x_2 - x_1$ into a larger number of parts, then $\sum_{n=0}^{N-1} f(x_1 + n\Delta x)\Delta x$ becomes a better approximation for the area under the curve $y = f(x)$ between x_1 and x_2 (Fig. 8.17). This suggests that if we let

$\Delta x \to 0$ by increasing N, the limit of the sum, i.e., the integral $\int_{x_1}^{x_2} f(x)dx$, will yield the exact area under the curve $y = f(x)$ between x_1 and x_2. In shorthand $\int_{x_1}^{x_2} f(x)dx = \text{area under } f(x) \Big|_{x_1}^{x_2}$.

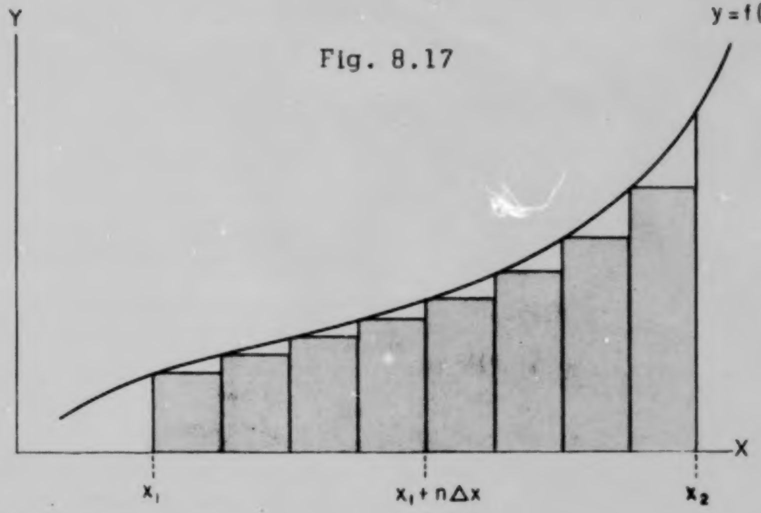


Fig. 8.17

We can, in fact, prove that this is the case. The sum of the areas of the rectangles in Fig. 8.17 is less than the exact area under the curve $y = f(x)$. However if we add the areas of the small rectangles shown by the broken lines in Fig. 8.18 to the areas of the rectangles of Fig. 8.17 we get an area that is larger than the area under the curve. Thus

$$\left| \text{Area under } f(x) \Big|_{x_1}^{x_2} - \sum_{n=0}^{N-1} f(x_1 + n\Delta x)\Delta x \right| < \text{sum of small rectangles}$$

All of the small rectangles have the base Δx and their heights add up to $f(x_2) - f(x_1)$. Thus the sum of their areas is $[f(x_2) - f(x_1)]\Delta x$. As $\Delta x \to 0$, $[f(x_2) - f(x_1)]\Delta x \to 0$ therefore

$$\left| \text{Area under } f(x) \Big|_{x_1}^{x_2} - \lim_{\Delta x \to 0} \sum_{n=0}^{N-1} f(x_1 + n\Delta x)\Delta x \right| \to 0$$

or

$$\text{Area under } f(x) \Big|_{x_1}^{x_2} = \int_{x_1}^{x_2} f(x)dx \qquad (12)$$

Rewriting Equation (11) as

$$\int_{x_1}^{x_2} f(x)dx = F(x_2) - F(x_1)$$

and combining it with Equation (12) we conclude that area under a curve $y = f(x)$ between x_1 and x_2 is given by the difference of values of an antiderivative of $f(x)$ at those points

$$\text{Area under } f(x) \Big|_{x_1}^{x_2} = F(x_2) - F(x_1)$$

It is not at all important which of the family of homomorphic antiderivative we choose, as long as we use the same one for both x_2 and x_1. Thus, the antiderivative provides a powerful tool for the calculations of areas under curves.

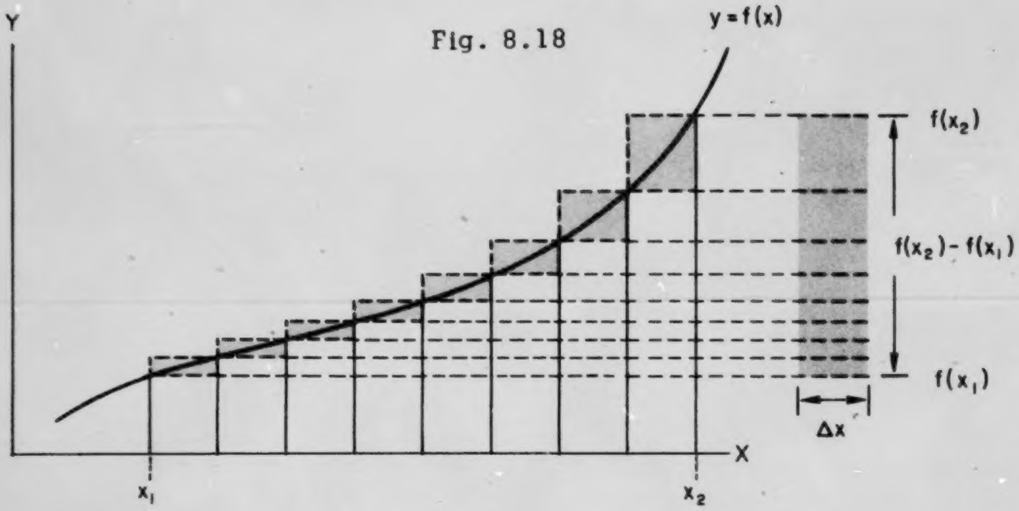


Fig. 8.18

Questions

1. The proof that $\int_{x_1}^{x_2} f(x)dx$ is equal to the area under the curve $y = f(x)$ between x_1 and x_2 makes use of the fact that the sum of the areas of the rectangles is less than the exact area under the curve. This is true only if the function $y = f(x)$ is increasing over the interval x_1 to x_2.

(a) Sketch a function $y = f(x)$ which decreases over an interval x_1 to x_2.

(b) Use rectangles touching the graph at their left corner to estimate the area under $y = f(x)$ from x_1 to x_2.

(c) Modify the proof of Section 8.10 to show that also in this case $\int_{x_1}^{x_2} f(x)dx = \text{area under } f(x) \Big|_{x_1}^{x_2}$.

2. (a) From the description of integrals as area under curves, show that

$$\int_{x_1}^{x_2} f(x)dx + \int_{x_2}^{x_3} f(x)dx = \int_{x_1}^{x_3} f(x)dx$$

(b) Using part (a) argue that $\int_{x_1}^{x_2} f(x)\, dx = \text{area under } f(x) \Big|_{x_1}^{x_2}$ even if $f(x)$ both increases and decreases over the interval x_1 to x_2.

3. Consider a function $f(x) > 0$ for $x_1 \leq x \leq x_2$. Let $g(x) = -f(x)$.

(a) How is the curve $y = g(x)$ related to the curve $y = f(x)$?

(b) If $F(x)$ is an antiderivative of $f(x)$, give an antiderivative of $g(x)$.

(c) How is the area "under" $y = g(x)$ related to the area under $y = f(x)$?

4. Find the area under the curve $y = x^2$ between x_1 and x_2.

5. What is the area under the curve $y = \frac{3}{x^2} + 5x - 2x^3$ in the interval $1 \leq x \leq 2$?

6. A function is called symmetric if $f(-x) = f(x)$ and antisymmetric if $f(-x) = -f(x)$. (For example, $f(x) = x^2$ is symmetric because $(-x)^2 = x^2$ and $f(x) = x^3$ is antisymmetric because $(-x)^3 = -x^3$.) What can you say about the area under $f(x) \Big|_{-x_1}^{x_1}$ where $f(x)$ is antisymmetric in the interval $-x_1 \leq x \leq +x_1$?

7. Under what condition will the area under a curve as calculated by Equation (12) be given in cm^2?

8.11 The Area Function

In Fig. 8.19 consider the point x_1 fixed and the point $\underline{X}$ moving along the x-axis. Then to every value $\underline{X}$ corresponds a value for the area under the curve between x_1 and $\underline{X}$. Or in other words, the area under a curve is a function of the upper end of the interval for a fixed lower end of the interval. We shall denote this function by A(X) called the area function of f(x).

$$\int_{x_1}^{X} f(x)dx = A(X) = \text{Area of } f(x)\Big|_{x_1}^{X} \qquad (13)$$

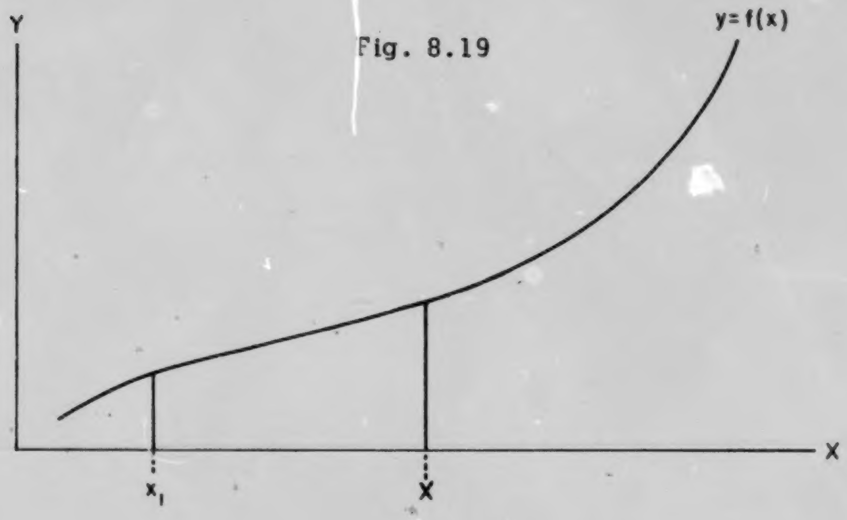


Fig. 8.19

Now compare Equation (13) for the antiderivative of f(x) and Equation (11) with x_2 replaced by $\underline{X}$:

$$F(X) = F(x_1) + \int_{x_1}^{X} f(x)dx \qquad (11')$$

Equation (11') reduces to Equation (13) for $F(x_1) = 0$. Thus the area function defined by Equation (13) is the particular antiderivative of f(x) that satisfies the initial condition $F(x_1) = 0$.

The connection between area function and antiderivative is very useful for getting a general feeling for the behavior of an antiderivative. For example, what are the qualitative features of the antiderivative of the function represented by the graph in Fig. 8.20 subject to the initial condition F(0) = 0?

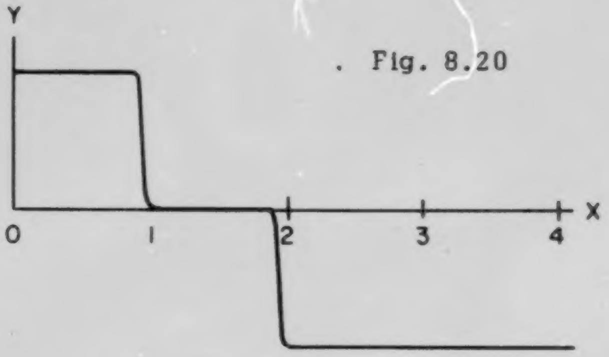


Fig. 8.20

Note, first of all, that for $0 < x < 1$ the area under the curve is proportional to x. Thus the antiderivative will start off as a straight line with a positive slope. Near $x = 1$ the function drops rapidly to zero and stays there up to $x = 2$. There is no change in the area under the curve in this interval, thus the antiderivative remains at its value at $x = 1$. The general appearance of the antiderivative between 0 and 2 is shown in Fig. 8.21.

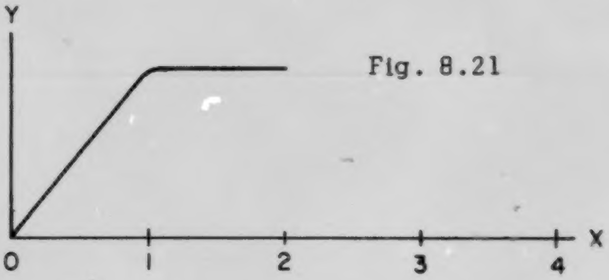


Fig. 8.21

Near $x = 2$ the function drops rapidly to minus its value between 0 and 1 and stays constant up to $x = 4$. Since $f(x)$ is negative and constant in the interval $2 < x < 4$ the area "under" the curve is negative and will reduce the area accumulated from $x = 0$ at a constant rate. The overall appearance of the antiderivative for $0 < x < 4$ is shown in Fig. 8.22.

A comment about notation is in order at this point. We have labeled the independent variable of the area function by X rather than x. The reason for this change is to avoid confusion between the upper end of the interval of integration and all the values inside it; when we write $\int_{x_1}^{X} f(x)dx$, x takes up all values between x_1 and X. Where there is no danger of confusion we can write F(x) for the antiderivative of f(x) without resorting to an X.

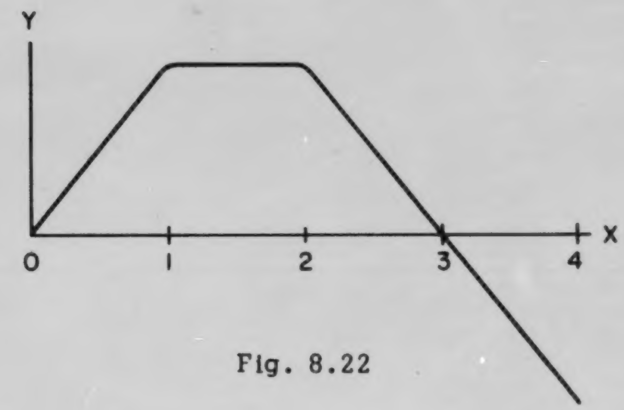


Fig. 8.22

Questions

1. It is stated in the text the area function of f(x) = c, where c is a constant, is a straight line. What is the slope of this line?
2. Sketch an antiderivative of the function f(x) described in Fig. 8.23. Assume F(-2) = 0. Check your answer by finding the algebraic expressions for the straight line segments and then calculating their antiderivatives.

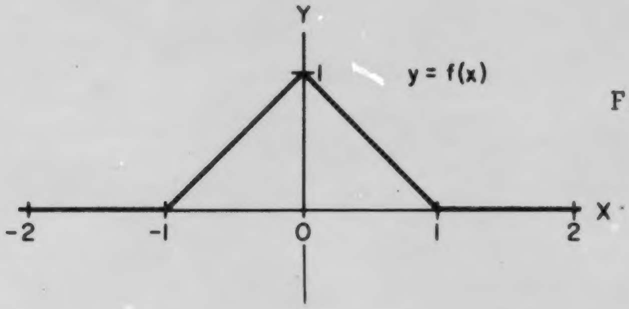


Fig. 8.23

3. By studying the area under the curve y = f(x) for

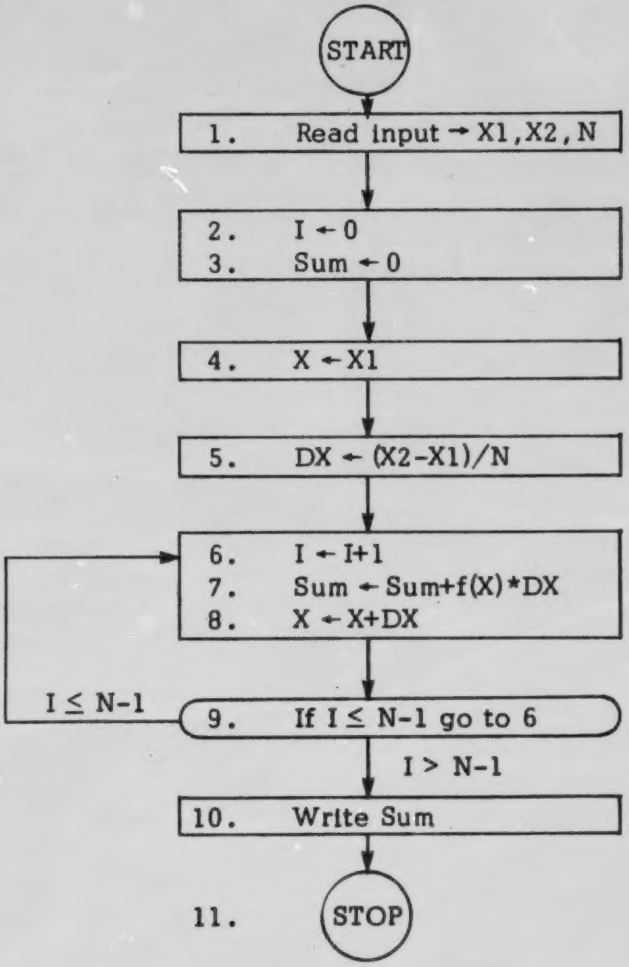
$$f(x) = \frac{1}{1 + x^2}$$

find the general features of its antiderivative subject to the initial condition F(0) = 0. What aspects of the behavior of F(x) can you deduce.

8.12 Numerical Integration

Computing the sum $\sum_{n=0}^{N-1} f(x_1 + n\Delta x)\Delta x$ as an approximation to $\int_{x_1}^{x_2} f(x)dx$ can lead to long and tedious calculations. Whenever possible, therefore, it is desirable to have this work done by a computer. A flow chart for a computer program which can be used to do this is shown as Program 1. The program requires that the values for x_1, x_2, and N be read in at Step 1. After initializing registers in Steps 2 through 4, the value for Δx, represented in the program by the register DX, is calculated in Step 5. Steps 6 through 8 calculate the actual sum, using register I as the index. Finally, the sum is printed in Step 10.

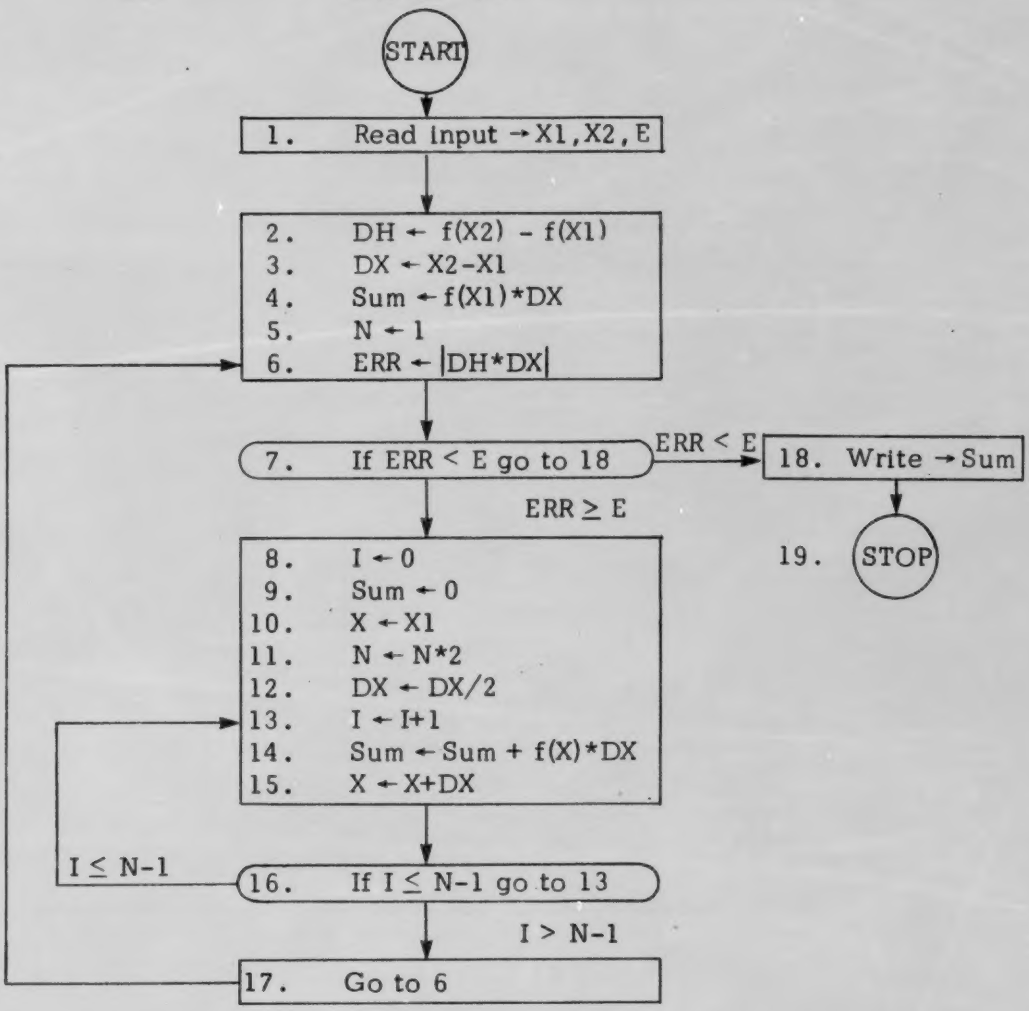
PROGRAM 1


START
1. Read input → X1,X2,N
2. I ← 0
3. Sum ← 0
4. X ← X1
5. DX ← (X2-X1)/N
6. I ← I+1
7. Sum ← Sum+f(X)*DX
8. X ← X+DX
I ≤ N-1
9. If I ≤ N-1 go to 6
I > N-1
10. Write Sum
11. STOP

A major limitation of this program is that it provides no information about the accuracy of the approximation to $\int_{x_1}^{x_2} f(x)dx$. To remedy this difficulty we can add a routine to check on the exactness of the approximation and, if necessary, improve it by increasing the number of subdivisions.

Program 2 is a version of the first program with these new features added in. The new program repeatedly doubles N until the approximation to $\int_{x_1}^{x_2} f(x)dx$ differs from the true value of $\int_{x_1}^{x_2} f(x)dx$ by less than some prespecified number E.

PROGRAM 2


START
1. Read input → X1,X2,E
2. DH ← f(X2) - f(X1)
3. DX ← X2-X1
4. Sum ← f(X1)*DX
5. N ← 1
6. ERR ← |DH*DX|
7. If ERR < E go to 18
ERR < E
18. Write → Sum
19. STOP
ERR ≥ E
8. I ← 0
9. Sum ← 0
10. X ← X1
11. N ← N*2
12. DX ← DX/2
13. I ← I+1
14. Sum ← Sum + f(X)*DX
15. X ← X+DX
I ≤ N-1
16. If I ≤ N-1 go to 13
I > N-1
17. Go to 6

Programs such as Program 1 and Program 2 require that the rule for $y = f(x)$ be specified. Therefore, if the function is presented either graphically or in tabular form such programs cannot be used. We can modify Program 1 to deal with tabular data but the approximation can no longer be made arbitrarily precise.

Questions

1. Code Program 1 in BASIC for $f(x) = x^2 + 4$. Use your program to approximate the area under $y = f(x)$ from $x_1 = 0$ to $x_2 = 2$ using four rectangles.
2. Modify Program 1 to approximate the area of a function given in tabular form. (Hint: You will have to read in values of x and f(x) instead of computing them.)
3. Code Program 2 in BASIC for $f(x) = x^2 + 4$. Use your program to approximate the area under $y = f(x)$ from $x_1 = 0$ to $x_2 = 2$ to within 0.01 unit squares.
4. (a) Modify Program 2 to print not only the final result but also the approximate value of $\int_{x_1}^{x_2} f(x)\,dx$ for each value of N.

   (b) Use your modified program to approximate the area under $f(x) = x^2 + 4$ from $x_1 = 0$ to $x_2 = 2$ to within 0.01 unit squares.

   (c) What does the series of successive approximations tell you about how such approximations are related to the true area?
5. Modify the program of Question 4 to compute various areas under $f(x) = \sqrt{1 - x^2}$ (see Fig. 8.24.)

   (a) First compute the area under $f(x) = \sqrt{1 - x^2}$ from $x_1 = -1$ to $x_2 = 0$. Use your result to approximate π to four decimal places.

   (b) Next, approximate π to four decimal places by computing the area under $f(x) = \sqrt{1 - x^2}$ from $x_1 = 0$ to $x_2 = 1$. How do the two estimates compare? One would expect the two methods to agree. Do they?

(c) Use your program to compute the area under $f(x) = \sqrt{1 - x^2}$ from $x_1 = -1$ to $x_2 = 1$. What happens? Why?

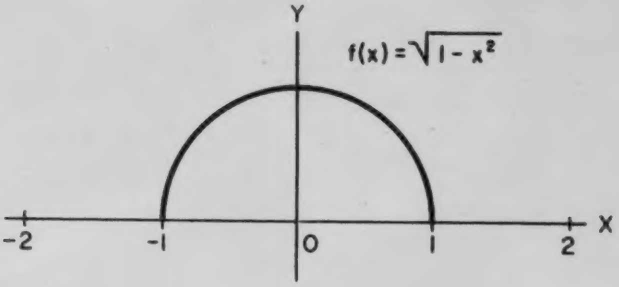


Fig. 8.24

6. (a) For Program 2 describe how the accuracy of the approximation is determined.

   (b) What are the limitations of the procedure used? (Hint: Consider the results of Problem 5.)

7. (a) Modify Program 2 so that the criterion for accuracy is based on the comparison of two consecutive approximations. That is, if two approximations differ by less than E the program should print the result and stop. Otherwise, divide the interval once again and continue the process.

   (b) How does this procedure compare to the one of Program 2?

   (c) Does this criterion work equally well for all integrals?

8. (a) Modify Program 2 so that it can be used to compute $A(X) = \int_{x_1}^{X} f(x)\,dx$ and print out a table of values of X and A(X) which could be used to graph the function A(X).

   (b) Use the program of part (a) to print a table for $A(X) = \int_{0}^{X} (x^2 + 4)\,dx$.

9. Use the program of Question 8 to print a table for $A(X) = \int_0^X \frac{dx}{1+x^2}$.

10. How could the program of Question 8 be modified to print out a table which could be used to plot $F(X) = F(x_1) + \int_{x_1}^X f(x)dx$?

Chapter 9. THE EXPONENTIAL AND LOGARITHMIC FUNCTIONS

9.1 The Exponential Function $y = 10^x$

In Chapter 4 you learned how to find intermediate points on a power-of-ten slide rule by taking square roots of ten. For example, the number on the D scale represented by a displacement from 1 to a point halfway between 1 and 10 is $\sqrt{10} = 10^{1/2} = 3.162$. By taking successive square roots you found still more intermediate points. Thus, to find the number on the D scale lying halfway between 1 and $10^{1/2}$ you calculated $\sqrt{10^{1/2}} = \left(10^{1/2}\right)^{1/2} = 10^{1/4}$.

By taking successive square roots of 10 you were, in fact, raising 10 to different fractional powers $10^{1/2}$, $10^{1/4}$, $10^{1/8}$, $10^{1/16}$ . . . . Each of the terms in the sequence of exponents $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, etc., has the form $\frac{1}{2n}$, where n is a positive integer. Using products of fractional powers of the form $10^{1/2n}$, we can find the value of any fractional power of ten to any desired accuracy. For example, if we wish to find the value of $10^{0.835}$, we first search for a sum made up of terms from the sequence of exponents $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$ that differs from the exponent 0.835 by only a little. The sum $\frac{1}{2} + \frac{1}{4} = 0.75$ is close to 0.835, but we can easily get closer to it by adding and subtracting additional well-chosen terms from the series. Thus,

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} = 0.875 \qquad \text{error: } 0.0400$$

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} - \frac{1}{32} = 0.8438 \qquad \text{error: } 0.0088$$

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} - \frac{1}{32} - \frac{1}{128} = 0.8359 \qquad \text{error: } 0.0009$$

To find the value of $10^{0.835}$ we now make use of the law of exponents, $a^m \cdot a^n = a^{m+n}$, and write

$$
\begin{aligned}
10^{0.835} &\approx 10^{\frac{1}{2}+\frac{1}{4}+\frac{1}{8}-\frac{1}{32}-\frac{1}{128}} \\
&= 10^{\frac{1}{2}} \cdot 10^{\frac{1}{4}} \cdot 10^{\frac{1}{8}} \cdot 10^{-\frac{1}{32}} \cdot 10^{-\frac{1}{128}} \\
&= \frac{10^{\frac{1}{2}} \cdot 10^{\frac{1}{4}} \cdot 10^{\frac{1}{8}}}{10^{\frac{1}{32}} \cdot 10^{\frac{1}{128}}}
\end{aligned}
$$

We can evaluate each of the fractional powers of ten in the above expression by taking successive square roots. If we do this (using a high-speed calculator to save time), we get for the final result

$$10^{0.835} \approx 6.85$$

Using this tedious but routine method, we can find the value of any fractional power of ten to any degree of accuracy. There are other ways of calculating fractional powers of ten, but this is the method invented and used (without the benefit of high-speed calculators) in the seventeenth century.

What we have just done is find the value of $f(x)$ in the function $f(x) = 10^x$ for $x = 0.835$. This new function is called an exponential function. The rule for this function is "take 10 to the x power," which is not hard to do for integral values of x, but, as you have just seen, is not so easy for many non-integral values of x. However, if we calculate a reasonable number of values of 10^x, we can draw a graph of this exponential function, filling in the hard-to-calculate gaps with a smooth curve.

The graph of the exponential function $y = 10^x$ is shown in Fig. 9.1. It rises steeply, passing through the point (0,1), and as x assumes larger and larger values, 10^x increases without limit, but as x becomes more and more negative, 10^x asymptotically approaches the x axis.

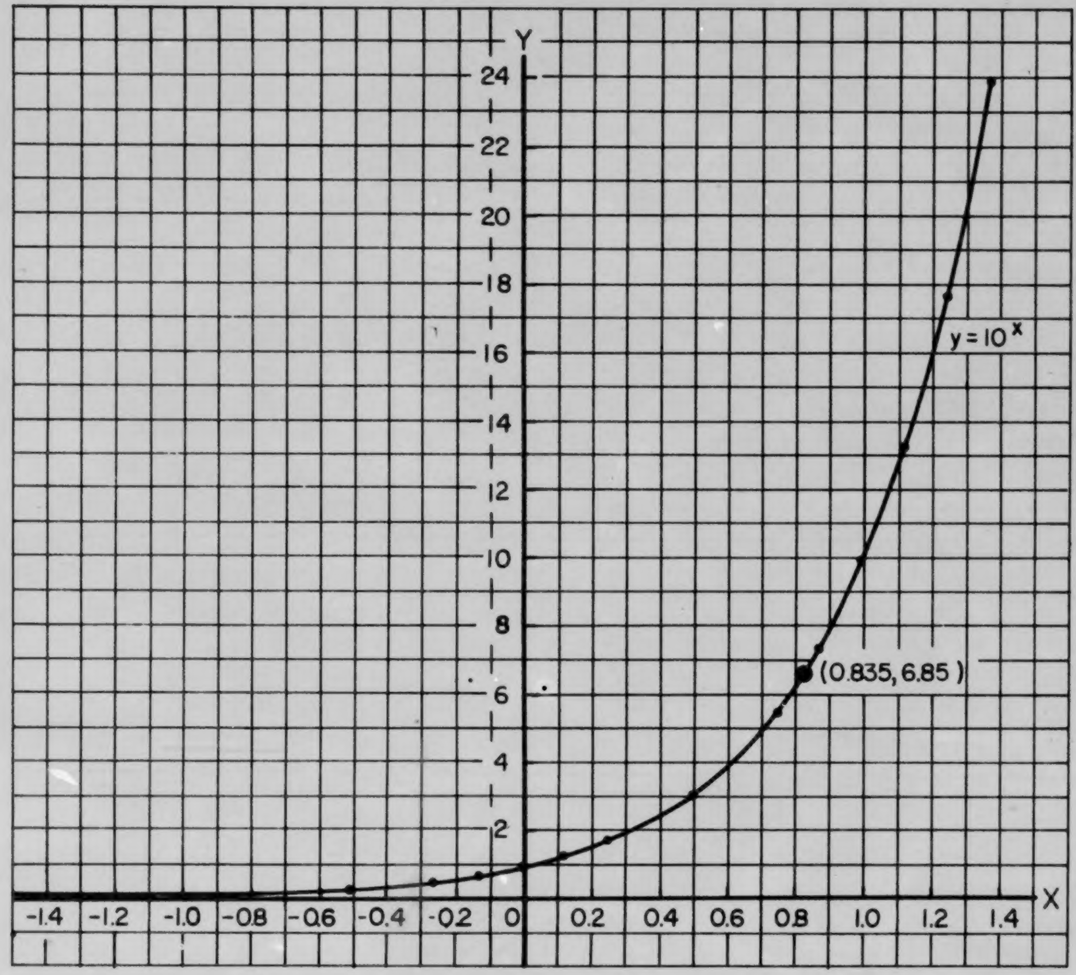


Fig. 9.1

Questions

1. Using the law of exponents, $a^m \cdot a^n = a^{m+n}$, show that $10^{1/2n}$ is the square root of $10^{1/n}$ where n = any integer.

2. Find, using successive square roots, the value of
   (a) $10^{1/4}$
   (b) $10^{0.125}$
   (c) $10^{2.5}$
   (d) $10^{1.125}$
   (e) $10^{-0.25}$

3. What are the domain and range of the exponential function $y = 10^x$?

4. Suppose you have a very long piece of string which you cut into 10 pieces, and you repeat this process of cutting each piece into 10 pieces several times.

(a) Make a graph of the number of pieces of string N as a function of the number of times n you repeat the process (start with $n = 0$ when $N = 1$).

(b) Are you justified in connecting the points you plot by a smooth curve?

(c) What function gives N as a function of n?

(d) What restriction must you apply to the independent variable of this function?

9.2 The Exponential Functions $y = b^x$ and $y = kb^x$

In the previous section we discussed the properties of the function $f(x) = 10^x$, which is a special case of the more general exponential function $f(x) = b^x$ where b is any positive number.

We found that we could calculate 10^x for any x to any accuracy we wished by calculating the product of some "well-chosen" successive square roots, starting with the square root of 10. In similar fashion, we can find the value of b^x for any positive b and any x by starting with the square root of b instead of the square root of 10. The domain of $y = b^x$ for any allowable choice of b extends over the whole number line. The range, however, consists of only positive values. The graph of $y = b^x$ for various choices of b is shown in Fig. 9.2. If b is greater than 1, then as x assumes larger and larger values, b^x increases without limit, but as x gets more and more negative, b^x asymptotically approaches the x axis. For values of b less than 1, the reverse holds true.

The exponential function is one which comes up very often. For example, let us say that the population of wild rabbits doubles each year. If we start with k rabbits, then after one year we have $2k$ rabbits, after two years $2(2k) = 2^2k$ rabbits, after three years $2(2^2k) = 2^3k$ rabbits, etc. After x years, by the same reasoning, there will be $y = (2^x)k$ rabbits. Here we have a more general form of the exponential function, namely

$$y = kb^x$$

The same function $y = kb^x$ describes the total amount of money in a savings bank account, assuming a constant rate of interest. Let us say that we start with a principal of m dollars and that the annual interest is 6 per cent, or 0.06 of the principal. After one year we have 1.06m dollars in the account, after two years $1.06 \times (1.06m)$, etc., and after x years we have $(1.06)^x m$ dollars. Of course, the balance shown in the account is not really a smoothly varying function of x, since interest is usually not credited continuously as it accrues, but is added as a lump sum at fixed intervals. Nevertheless, if the total time considered is large compared to the time between

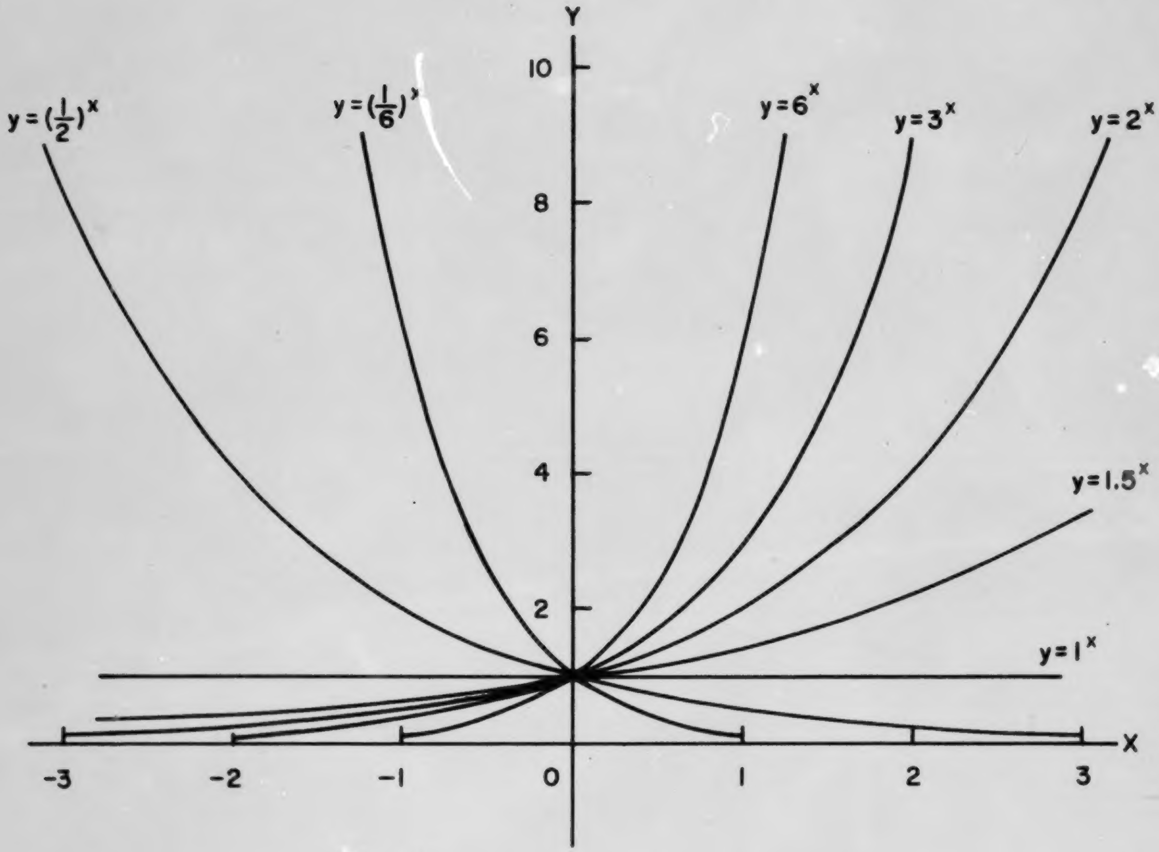
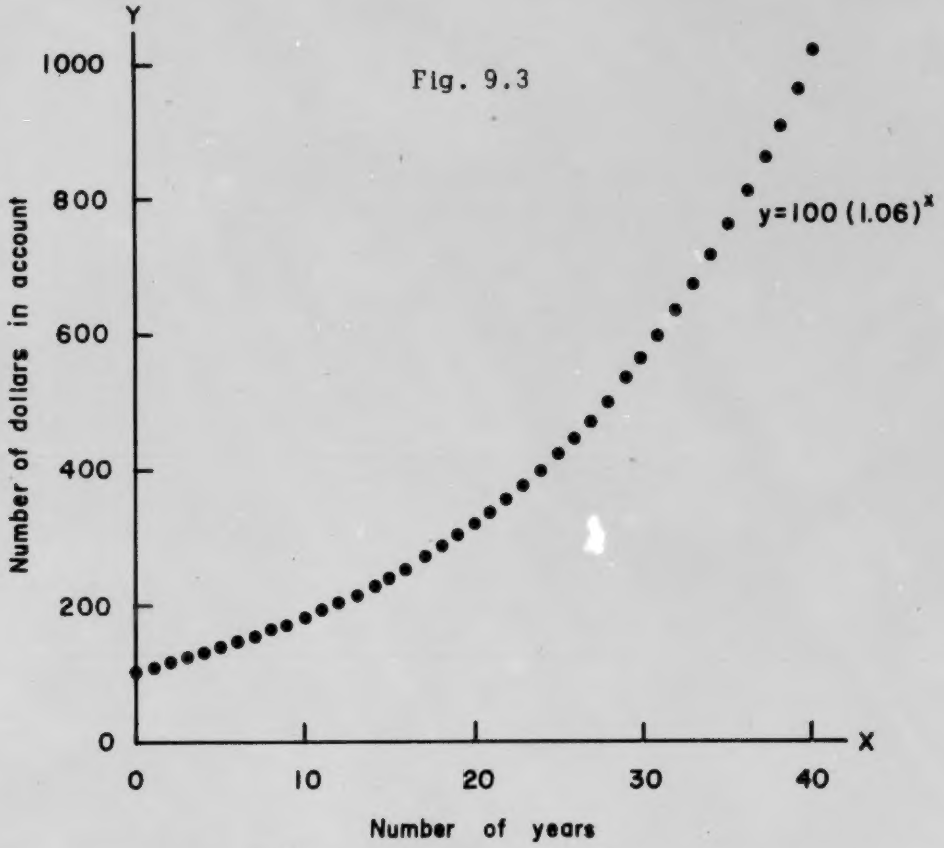


Fig. 9.2

interest dates, the function can for all practical purposes be considered smooth. (See Fig. 9.3.)



Questions

1. The exponential function $y = kb^{x-x_0}$, where k, b, and x_0 are constants, reduces to the form $y = k'b^x$, where k' is a constant. What is k' in terms of k, b, and x_0?
2. The exponential function $y = kb^{ax}$, where k, b, and a are constants, reduces to the form $y = k'c^x$, where c is a constant. What is c in terms of b and a?

9.3 Inverse Functions

In Section 6.1 we defined a function as a relation such that for each value of the independent variable there is only one value of the dependent variable. We said that we could look at a function graphically, in tabular form, or we could express a function in terms of a rule.

It is sometimes useful to think of a function as a rule by which we pair off certain numbers with other numbers; we may consider the function $\underline{f}$ as a rule that pairs off a number $\underline{x}$ with the number $f(x)$. Certain of these rules can be inverted; that is, another function can be found converting $f(x)$ back into $\underline{x}$.

For example, the function $f(x) = x + 3$ pairs off 0 with $f(0) = 3$, 1 with $f(1) = 4$, 10 with $f(10) = 13$, etc.. The function $g(x) = x - 3$ inverts the rule of $f(x) = x + 3$ since it pairs off 3 with $g(3) = 0$, 4 with $g(4) = 1$, 13 with $g(13) = 10$, etc..

It is not always possible to find a function inverting the rule of another function. For example, we cannot find a function that is the inverse of the function $f(x) = x^2$. The reason is that $f(2) = 4$ and $f(-2) = 4$, so a rule $g(x)$ which inverts the rule $f(x)$ would have to satisfy both $g(4) = 2$ and $g(4) = -2$. But if the rule $g(x)$ defines a function, then $g(4)$ must be a unique number; thus there is no function which is the inverse of $f(x) = x^2$.

Whenever we have a function $g(x)$ which reverses the rule of a function $f(x)$, then $g(x)$ is called the inverse function of $f(x)$.

What is the graphical relationship between a function and its inverse? Let us first make the observation (Fig. 9.4) that the line $y = x$ is the perpendicular bisector of the line segments connecting the points (a,b) and (b,a), (c,d) and (d,c), (m,n) and (n,m). Or, to rephrase the statement, the points (a,b), (c,d), and (m,n) are symmetric to the points (b,a), (d,c), and (n,m), respectively, about the line $y = x$.

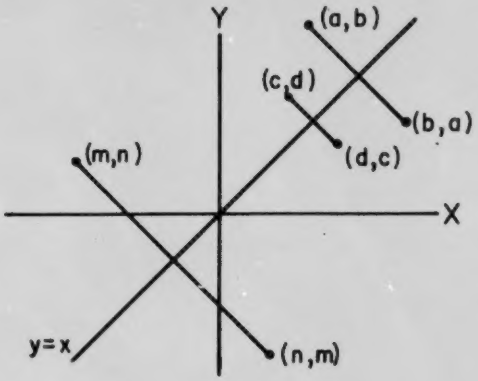


Fig. 9.4

We know that to graph a function f(x), with values for f(a) = b, f(c) = d, f(m) = n, etc., we plot the points (a,b), (c,d), (m,n), etc. Since the inverse reverses the rule of the function, we can graph the inverse by plotting the points (b,a), (d,c), (n,m), etc. In view of our observation concerning Fig. 9.4, it is now clear that the graph of the inverse of a function is symmetric to the graph of the function with respect to the line $y = x$.

Thus, for example, to sketch the graph of the inverse of f(x) in Fig. 9.5, even though we have no explicit rule that defines f(x), we can select a few points (P_1, P_2, ..., P_7) on the graph of f(x) and locate points of symmetry (Q_1, Q_2, ..., Q_7) with respect to the line $y = x$. We then sketch the graph of the inverse by connecting these points. We can tell by looking at the graph in Fig. 9.5 that the inverse is a function even though we cannot write the rule for it in terms of algebraic operations.

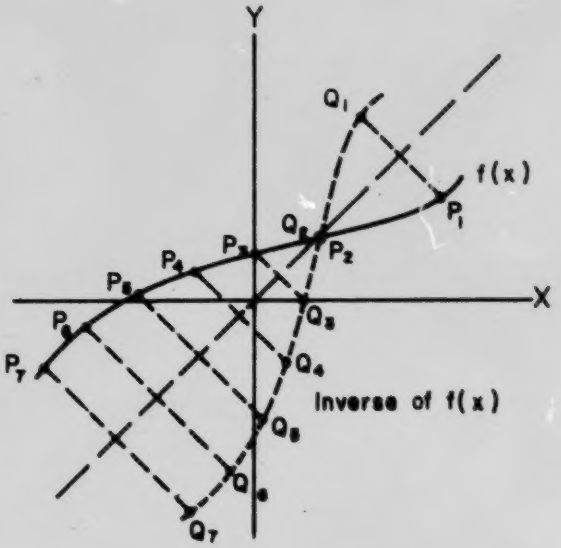


Fig. 9.5

We stated earlier in this section that the function $y = x^2$ does not have an inverse function. We can illustrate this graphically (Fig. 9.6(a)). Note that the relation $y = \pm\sqrt{x}$ reverses the rule $y = x^2$ but it is not a function since for any x in the domain of $y = \pm\sqrt{x}$ there are two values of the dependent variable y. It is possible, however, to get an inverse function for $f(x) = x^2$ if we restrict the domain of $f(x)$ to non-negative values only. The rule $y = +\sqrt{x}$ then defines the inverse function of $y = x^2$ where the domain of the independent variable in the function $y = x^2$ has been restricted to non-negative values of x. This is shown graphically in Fig. 9.6(b) where the solid portion of each graph shows the function and its inverse function.* The dotted portions are included just to show the complete relation between x and y.

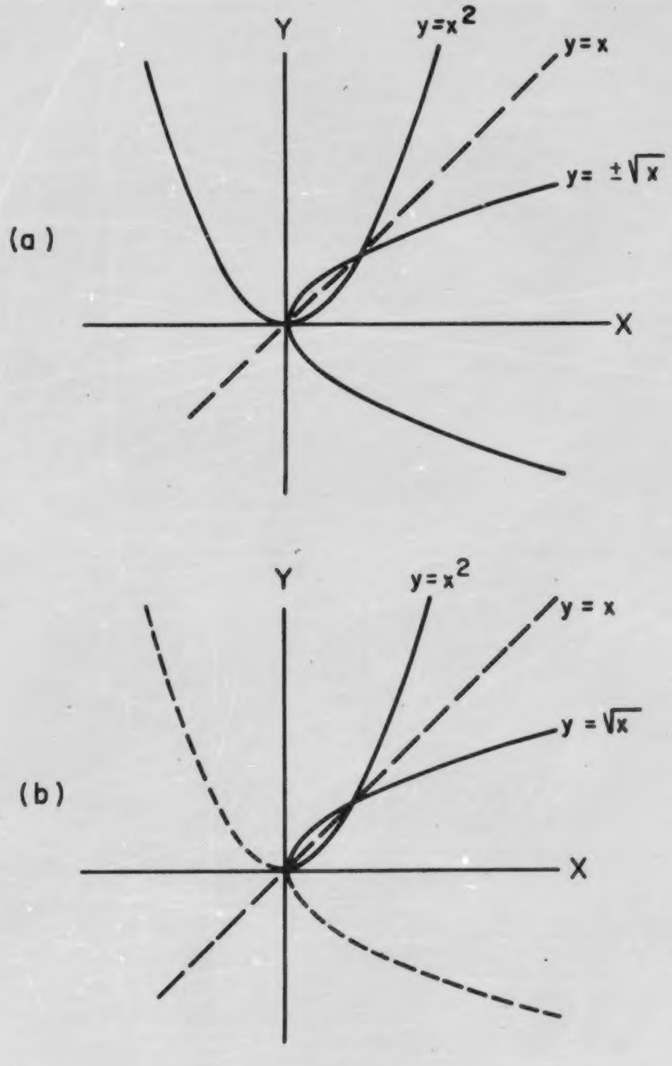


Fig. 9.6

---

[*] The function $y = +\sqrt{x}$ is usually written as $y = \sqrt{x}$, where the positive sign is understood.

Questions

1. (a) Figure 9.7(a) shows the function f(x) = 2x + 1 and its inverse g(x).

What is the algebraic expression for g(x)?

(b) Figure 9.7(b) shows f(x) = 2x + 1 and y = x. In this figure the scale on the x-axis has been "stretched" so that the distance from the origin to 1 along the x-axis is twice the corresponding distance along the y-axis.

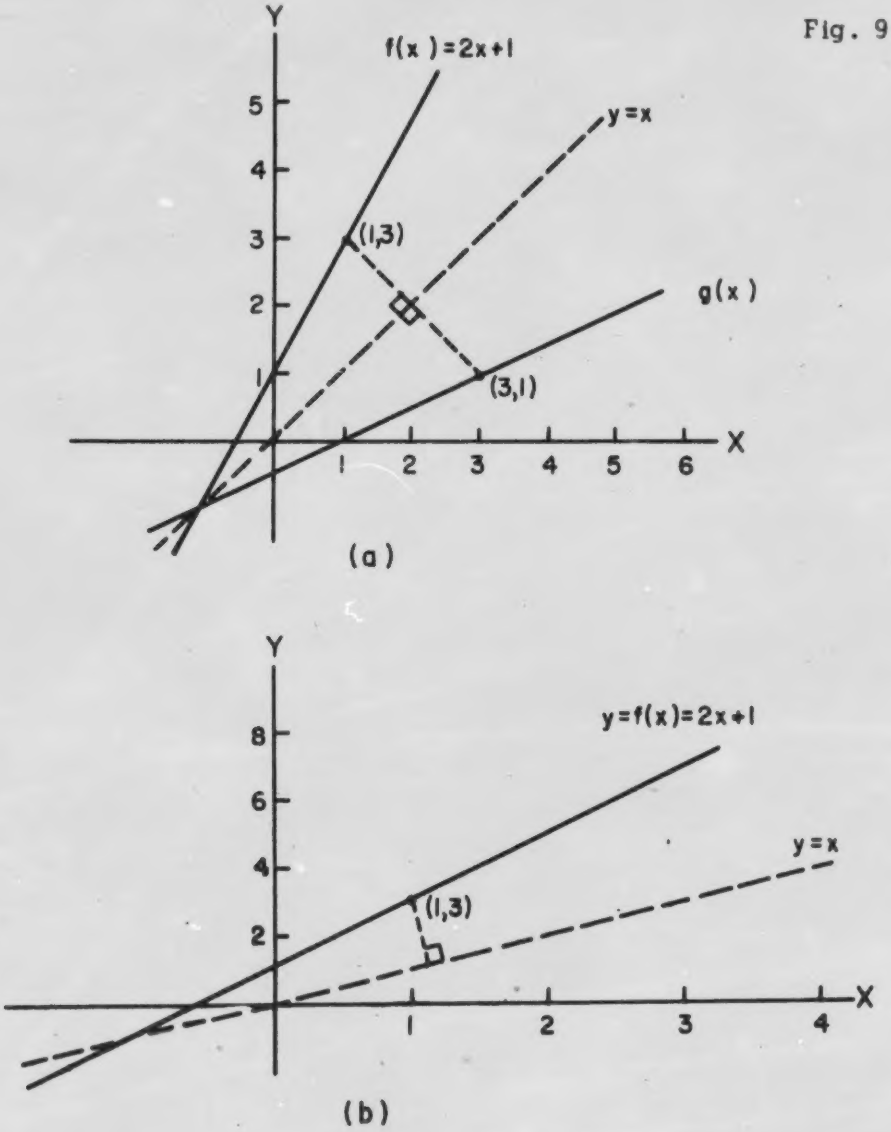


Fig. 9.7

Lay off, on graph paper, scales like those in Fig. 9.7(b) and plot the graph of $f(x) = 2x + 1$ and $y = x$.

Now use the algebraic expression you obtained in part (a) to plot $g(x)$.

(c) Are $f(x)$ and $g(x)$ symmetric with respect to the line $y = x$?

(d) Under what conditions are a function and its inverse symmetric with respect to the line $y = x$?

2. Is $y = -\sqrt{x}$ the inverse function of a function? If so, what function?

3. Sketch the graph of the inverse of each function in Fig. 9.8 without writing any algebraic expressions but by using the graphical relationships between a function and its inverse.

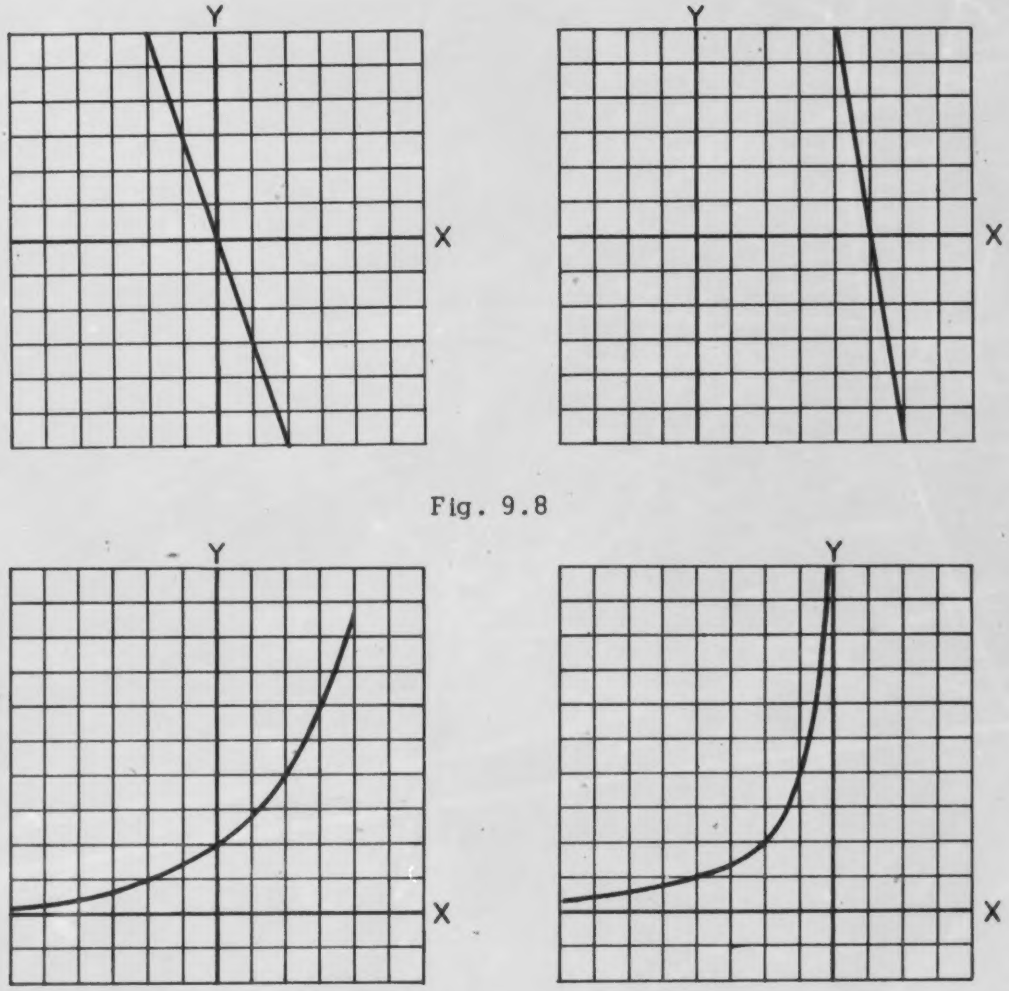


Fig. 9.8

4. In Fig. 9.9, the functions $f(x) = x^2$, $g(x) = 2$, and $h(x) = \sqrt{1 - x^2}$ are graphed.

   (a) Sketch the graph of the inverse of each function.

   (b) Is the inverse of any of these functions a function?

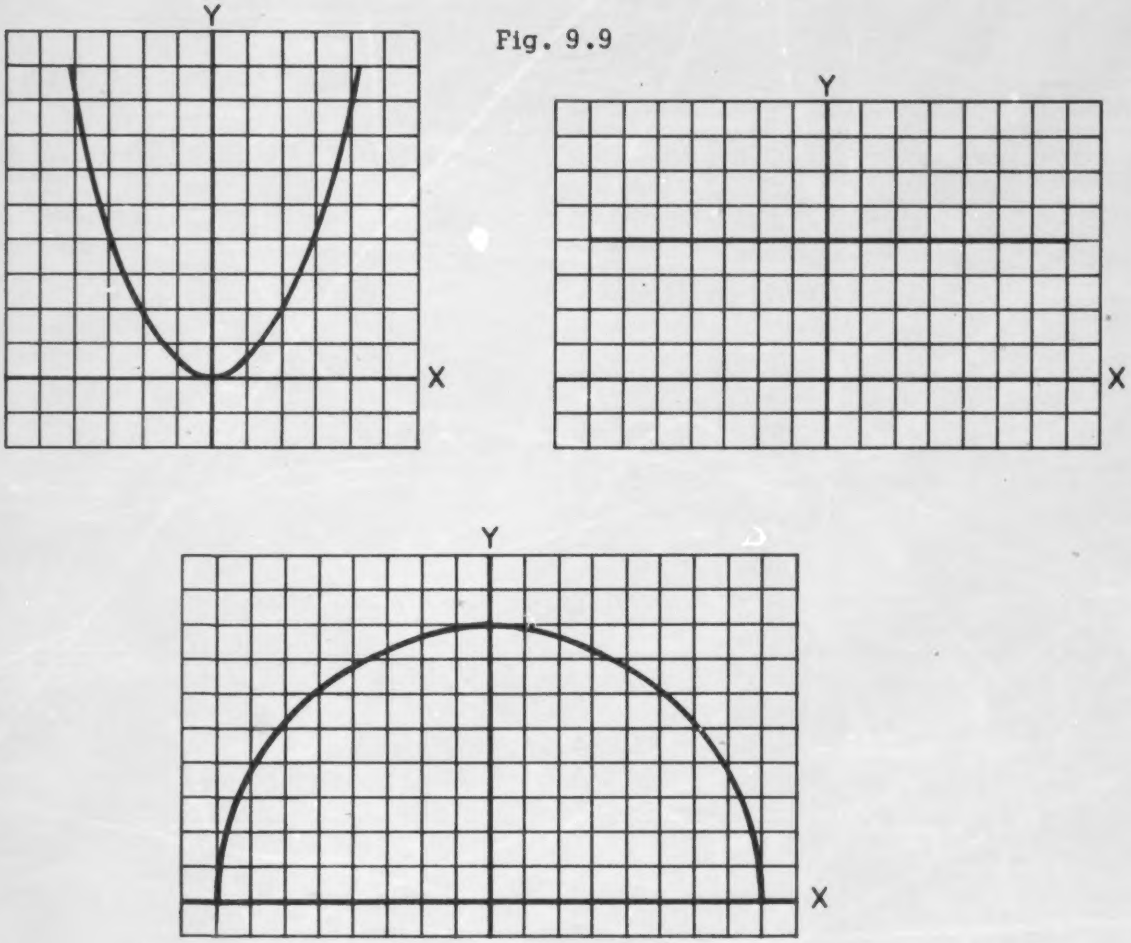


Fig. 9.9

5. Given a graph of a function, formulate a rule which will tell you, without going through the graphical construction of an inverse, whether or not the inverse is also a function?

6. Graph the function $y = f(x) = 4 - \frac{4}{3}x$ and then choose a few points on the graph of f(x) to plot its inverse. Now, using the technique described in Section 7.5, write the equation of its inverse.

7. In Question 6 you graphed the function $y = 4 - \frac{4}{3}x$. Solve the equation $y = 4 - \frac{4}{3}x$ for $\underline{x}$, and compare your solution with the equation of the inverse function that you derived in that example. Does this give you a method of writing the equations for inverses of linear functions? Try some more examples.

9.4 $g(x) = \log_{10}x$: The Inverse Function of $f(x) = 10^x$

Figure 9.10 shows the result of applying the geometric method for constructing the graph of an inverse of a function to the function $f(x) = 10^x$.

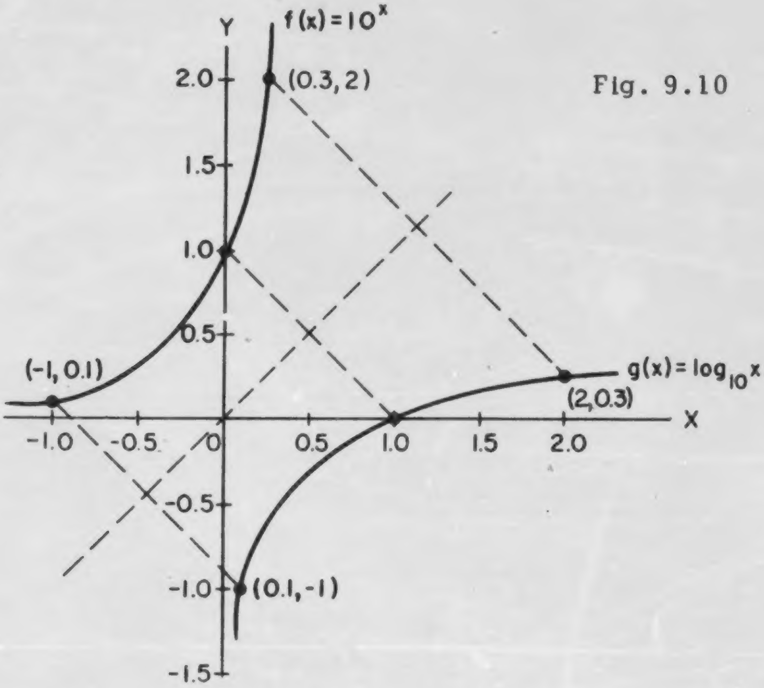


Fig. 9.10

As you can see, the inverse of $f(x) = 10^x$ is a function because for each value of the independent variable there is only one value of the dependent variable. It is called a logarithmic function. There is no way to write the exact rule for $g(x)$ using simple algebraic symbols, although we can find $g(x)$ for any positive value of $\underline{x}$ from a table of values of the exponential function $y = 10^x$. Therefore, we write it as

$$g(x) = \log_{10}x$$

where $\log_{10}x$ means "find the exponent of 10 such that $10^{g(x)} = x$" and is read as "the logarithm of $\underline{x}$ to base 10." Note that "$\log_{10}$" does not represent a number. Like the symbol "$\sqrt{\ }$" it specifies a definite operation.

From the definition of the logarithmic function it follows that

$$\begin{aligned} \log_{10}1 &= 0 & \text{since} \quad 10^0 &= 1 \\ \log_{10}10 &= 1 & \text{since} \quad 10^1 &= 10 \\ \log_{10}100 &= 2 & \text{since} \quad 10^2 &= 100 \\ &\vdots & &\vdots \end{aligned}$$

and for numbers less than 1

$$\begin{aligned} \log_{10}0.1 &= -1 & \text{since} \quad 10^{-1} &= 0.1 \\ \log_{10}0.01 &= -2 & \text{since} \quad 10^{-2} &= 0.01 \\ &\vdots & &\vdots \end{aligned}$$

Table 9.1 lists some of the overall properties of the two graphs in Fig. 9.10 and shows the close relationship between the functions $y = 10^x$ and $y = \log_{10}x$. We see that one graph behaves in just the reverse way from the other.

TABLE 9.1

	$f(x) = 10^x$	$g(x) = \log_{10}x$
Domain	all numbers on the number line	positive numbers on the number line
Range	positive numbers on the number line	all numbers on the number line
Intercept	(0, 1) with the y-axis	(1, 0) with the x-axis
Asymptote	x-axis	y-axis

The most characteristic property of the exponential function is expressed by the law of exponents. Specifically for base 10,

$$10^{x_1} \cdot 10^{x_2} = 10^{x_1+x_2} \tag{1}$$

What is the corresponding property for the logarithmic function?

Let

$$y_1 = 10^{x_1} \quad \text{and} \quad y_2 = 10^{x_2}$$

then by the definition of the logarithmic function*

$$x_1 = \log y_1 \quad \text{and} \quad x_2 = \log y_2$$

Equation (1) can now be written as

$$y_1 y_2 = 10^{x_1 + x_2}$$

Again applying the definition of the logarithmic function we have

$$x_1 + x_2 = \log (y_1 y_2)$$

On the other hand

$$x_1 + x_2 = \log y_1 + \log y_2$$

Hence

$$\log (y_1 y_2) = \log y_1 + \log y_2 \tag{2}$$

In words, the logarithm of the product of two numbers equals the sum of the logarithms of the two numbers.

Because of this relationship, a table of logarithms of numbers need include only the logarithms of numbers between 1 and 10. The table of logarithms to the base 10 in the Appendix of this book, for example, gives the logarithms of all three-digit numbers from 1 to 10 only, and the logarithms of these numbers, given to four digits, range from 0 to 1. (For simplicity, the decimal points in the numbers and also in the logarithms are omitted.)

To use such a table for numbers greater than 10 or less than 1, we express the number as a number between 1 and 10 multiplied by the appropriate power of 10. We then use Equation (2). For example

$$\begin{aligned} \log 324 &= \log (3.24 \times 10^2) \\ &= \log 3.24 + \log 10^2 \\ &= 0.5105 + 2 \\ &= 2.5105 \end{aligned}$$

---

*The subscript 10 is generally omitted when we write logarithms to the base 10. Thus, $\log y$ stands for $\log_{10} y$.

Similarly

$$\begin{aligned}\log 0.0324 &= \log(3.24 \times 10^{-2})\\ &= \log 3.24 + \log 10^{-2}\\ &= 0.5105 - 2\\ &= -1.4895\end{aligned}$$

Questions

1. The law of exponents holds for any number of factors: For example, $10^{x_1} \cdot 10^{x_2} \cdot 10^{x_3} = 10^{x_1+x_2+x_3}$.

   (a) Use this extention to show that

   $$\log y_1 y_2 y_3 = \log y_1 + \log y_2 + \log y_3$$

   (b) Express this result in compact form for the special case $y_1 = y_2 = y_3$.

   (c) How would you generalize this result to any number of equal factors?

2. In Equation (2) consider the special case $y_2 = \frac{1}{y_1}$. What does the result tell you about $\log \frac{1}{y_1}$ in terms of $\log y_1$?

3. The logarithm of the quotient $\frac{a}{b}$ can be looked upon as the logarithm of the product $a\frac{1}{b}$. On the basis of your answer to the preceding question state in a sentence the relationship between the logarithm of a quotient and the logarithms of its numerator and denominator.

4. How is $\log(x^{-n})$ related to $\log x$?

5. How is $\log\sqrt{x}$ related to log to $\log x$?

6. (a) Graph $y = \log x$ for values of x between 100 and 1000.

   (b) If you changed the x scale so that 100 became 1000, and 1000 became 10,000, how would you have to change the labeling of the y axis so that the graph would represent $\log x$ in the new domain?

7. Expand or simplify

   (a) $\log(ax^2)$

   (b) $\log(\frac{a}{x^2})$

8. Use the table of logarithms in the Appendix to find the logarithms of

the following numbers:

(a) 372

(b) 0.50

(c) 0.00437

(d) 3.46×10^8

(e) $\frac{1}{367}$

(f) $\frac{1}{0.021}$

9. Use the table of logarithms to evaluate the quantities below:

(a) $(1.72)^{18}$

(b) $(2.63)^{1/3}$

(c) $(143)^{-8}$

9.5 The Functions e^x and $\ln x$

Figure 9.11(a) shows two curves of the form $y = b^x$. They are $y = 2^x$ and $y = 4^x$. Both curves have the same y-intercept, (0,1), and both are approximately straight lines for $|x| << 1$. They can therefore be approximated

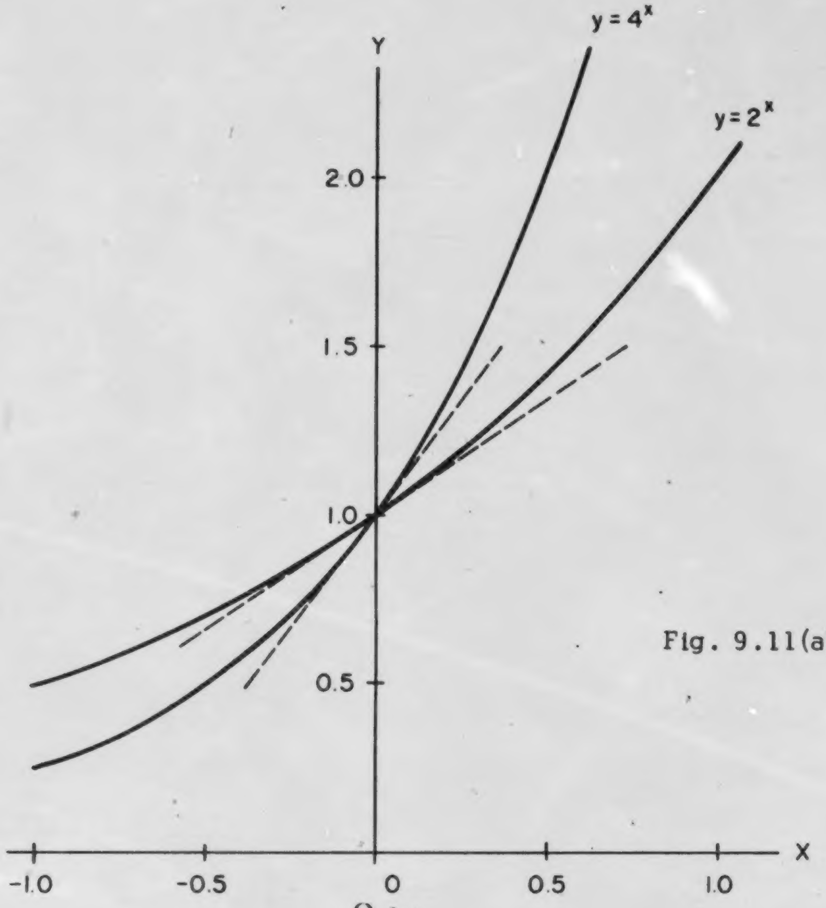


Fig. 9.11(a)

by the function $y \approx 1 + ax$ for $|x| << 1$ where a is the slope. As you can see from Fig. 9.11(a), at $x = 0$, $y = 4^x$ has a slope greater than 1, and $y = 2^x$ has a slope less than 1. There must exist an exponential function, $y = b^x$, whose slope at $x = 0$ is $a = 1$. This exponential function will have the simple approximation $y \approx 1 + x$. The base of this function is called e and the graph of $y = e^x$ and $y = 1 + x$, close to $x = 0$, is shown in Fig. 9.11(b). The figure shows that, indeed, for $|x| << 1$, the function $1 + x$ is a good approximation for e^x, so we write

$$e^x \approx 1 + x \qquad |x| << 1$$

We can find the value of the base e by taking both sides of the above equation to the $\frac{1}{x}$ power:

$$(e^x)^{1/x} = e \approx (1 + x)^{1/x} \qquad |x| << 1$$

Since our approximation $e^x \approx 1 + x$ becomes better as x approaches zero, we expect that the approximation $e \approx (1 + x)^{1/x}$ becomes better as x approaches zero. Table 9.2 gives the calculated value of $(1 + x)^{1/x}$ for a range of values of x approaching zero.

TABLE 9.2

x	$(1 + x)^{1/x}$
1	2.000
10^{-1}	2.594
10^{-2}	2.705
10^{-3}	2.717
10^{-4}	2.718
10^{-5}	2.718
10^{-6}	2.718

The table shows that for $|x| < 10^{-4}$ there is no change, to four significant digits, in the values of $(1 + x)^{1/x}$. We can say, therefore, that $e = 2.718$ to four significant digits.

Exponential functions with the base e occur frequently and for this reason tables of e^x and e^{-x} are found in many textbooks and handbooks.

The inverse of the function $y = e^x$ is also common. This logarithmic function $y = \log_e x$, is commonly written as $y = \ln x$ to distinguish it from the only other commonly used logarithmic function, $y = \log x$.*

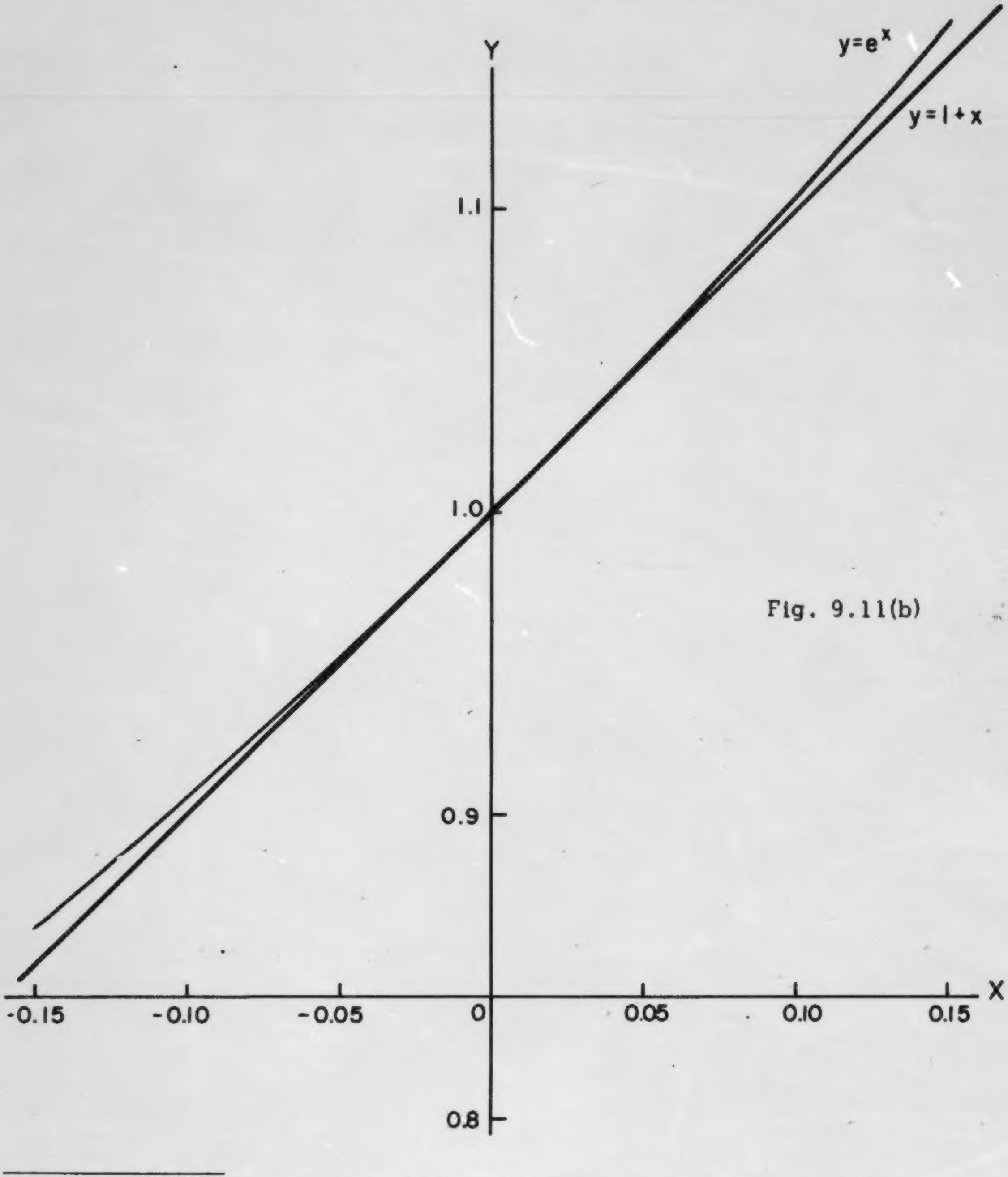


Fig. 9.11(b)

---

*Logarithms to the base e are called "natural logarithms," or sometimes "Naperian logarithms" after their inventor. (Logarithms to the base 10 are often called "common logarithms.")

What is the relationship between $\log x$ and $\ln x$? To find it, we take the logarithm to the base e of x for the case where $x = 10^y$. Thus $\ln x = y \ln 10$ or $y = \frac{\ln x}{\ln 10}$ and $\log x = y \log 10 = y$. Thus

$$\log x = \frac{1}{\ln 10} \ln x$$

Since $\frac{1}{\ln 10}$ is a constant, we see that logarithms to the base 10 are proportional to logarithms to the base e.

Another important property of $\ln x$ is that we can readily derive an approximation for $\ln(1 + x)$ for $|x| << 1$ This is a direct result of the approximation $e^x \approx 1 + x$. Thus, by definition of the logarithmic function,

$$\ln(1 + x) \approx x \qquad |x| << 1$$

One should not forget that this approximation applies only to logarithms to the base e. It does not apply to logarithms to the base 10.

Questions

1. What is the relationship between e^x and 10^x?
2. What is the value of
   (a) e^{10}
   (b) e^{100}
3. Use a table of e^x to find the fractional error in the approximation $e^x \approx 1 + x$ for
   (a) $x = 0.01$
   (b) $x = 0.1$
   (c) $x = 0.5$
4. The expression $\log x = \frac{1}{\ln 10} \ln x$ makes it possible to obtain $\log x$ from tables of $\ln x$. What is the corresponding expression that allows us to obtain $\ln x$ from tables of $\log x$?
5. Using the table of logarithms in the Appendix find
   (a) $\ln 1$
   (b) $\ln 10$
   (c) $\ln 100$

6. What is the fractional error in the approximation $\ln(1 + x) \approx x$ for
   (a) $x = 0.001$
   (b) $x = 0.1$
   (c) $x = 0.5$

7. If a sum of money increases by a fixed, small percentage at regular time intervals, then the amount A at time t in years is given by

$$A = m\left(1 + \frac{r}{n}\right)^{nt}$$

where m is the amount when t is zero, r is the interest rate and n is the number of times per year the interest is added to the principal (compounded).

(a) Express $\left(1 + \frac{r}{n}\right)^{nt}$ as e^x.

(b) By making use of the approximation $\ln(1 + x) \approx x$ for $x << 1$ find the expression for A when the interest is compounded continuously $(n \to \infty)$.

(c) What is the difference between \$1000 compounded annually at 6 per cent and \$1000 compounded continuously at the same rate?

9.6 The Derivative of $y = e^x$; Exponential Growth and Decay

The exponential function has many applications. To study these applications we need a knowledge of its derivative.

According to Equation (2) of Section 8.3, the derivative of the function $f(x) = e^x$ is defined as

$$[e^x]' = \lim_{\Delta x \to 0} \frac{e^{x+\Delta x} - e^x}{\Delta x}$$

or

$$[e^x]' = \lim_{\Delta x \to 0} \frac{e^x(e^{\Delta x} - 1)}{\Delta x} \qquad (3)$$

Unfortunately we cannot write the numerator in Equation (3) in such a way that Δx in the numerator can be cancelled with the Δx in the denominator. However we know from Section 9.5 that $e^{\Delta x} \approx 1 + \Delta x$ when $\Delta x << 1$

and that this approximation approaches an equality in the limit as $\Delta x \to 0$. Therefore Equation (3) becomes

$$[e^x]' = \lim_{\Delta x \to 0} \frac{e^x(1 + \Delta x - 1)}{\Delta x} = e^x \tag{4}$$

The exponential function (with base *e*) has the interesting property that it equals its own derivative!

How is this property modified for the more general exponential function e^{kx}? Again, applying the delta process we find

$$[e^{kx}]' = \lim_{x\Delta \to 0} \frac{e^{k(x+\Delta x)} - e^{kx}}{\Delta x} = \lim_{\Delta x \to 0} \frac{e^{kx}(e^{k\Delta x} - 1)}{\Delta x}$$

For any value of *k* we can choose Δx so small that also $k\Delta x$ will fulfill the condition $k\Delta x \ll 1$. Then we apply the approximation from Section 9.5:

$$e^{k\Delta x} \approx 1 + k\Delta x$$

and find

$$[e^{kx}]' = \lim_{\Delta x \to \infty} \frac{e^{kx}(1 + k\Delta x - 1)}{\Delta x} = ke^{kx}$$

In words, the derivative of e^{kx} is proportional to the function itself and the constant of proportionality is *k*.

What is the derivative of Ae^{kx}? From theorem (i) of Section 8.5, we know that the derivative of a constant times a function is the constant times the derivative of that function. Hence

$$[Ae^{kx}]' = A[e^{kx}]' = Ake^{kx} = kAe^{kx} \tag{5}$$

Thus the derivative of Ae^{kx} is proportional to Ae^{kx} itself and the constant of proportionality *k* is independent of the value of *A*. The converse of this result is also true. We state it here without proof:

Any function $f(x)$ which has the property that its derivative is proportional to itself, is an exponential function. Specifically, if

$$f'(x) = kf(x)$$

then

$$f(x) = Ae^{kx}$$

Note that $f(0) = A$. Hence, A can be specified by an initial condition. For example, if $f'(x) = 1.5f(x)$ and $f(0) = 10$, then $f(x) = 10e^{1.5x}$.

In many applications of exponential functions the independent variable is time and dependent variable is the number of such discrete things as atoms, bacteria, people, etc. In such cases we speak of population functions, and designate the dependent variable by $\underline{N}$.

Population functions change by discrete amounts and therefore have the property that their graphs are not smooth curves. But if over a short time interval, the changes in the size of the population are small compared to the total population considered, then for all practical purposes we can consider a population function to be smooth and speak of a rate of change N'.

In this notation, if a population function satisfies the equation,

$$N' = kN$$

then it is of the form

$$N = N_0e^{kt} \tag{6}$$

where N_0 is the size of the population at $t = 0$.

Questions

1. Find the derivative of each of the following exponential functions:
   - (a) $3e^{x}$
   - (b) $4e^{-x}$
   - (c) $0.5e^{3x}$
   - (d) $5e^{-0.1x}$
2. Evaluate each of the derivatives in Question 1 for $x = 2$.
3. For which of the functions in Question 1 does the value of the function increase with increasing values of $\underline{x}$?
4. Which of the following functions would probably be of exponential form?

   (a) $s(n)$, your annual salary $\underline{n}$ years after beginning your job if you have been promised annual pay raises of 5 per cent.

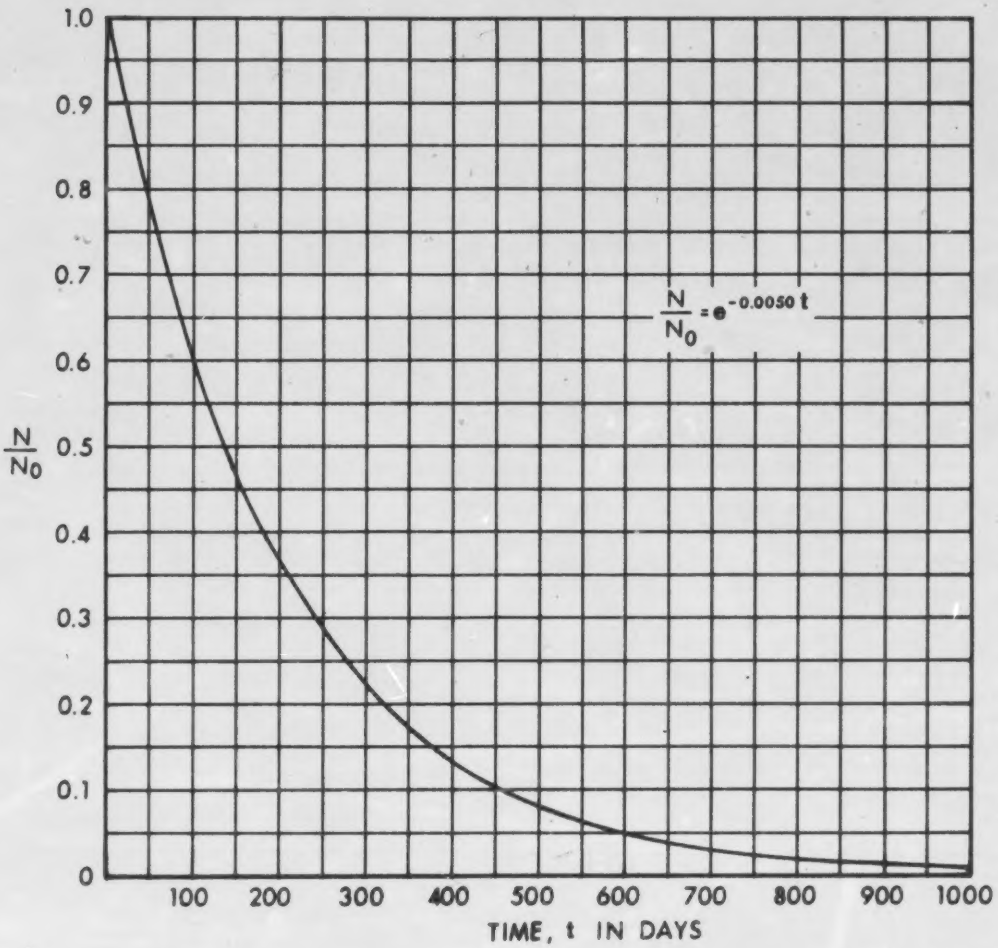
(b) f(t), the temperature on a hot day at time t.

(c) n(t), the number of people who have heard a rumor t days after it was started.

(d) f(n), your annual salary n years after beginning your job if you have been promised annual pay raises of $700.

5. Figure 9.12 is an illustration of Equation (5). It is the graph of the decay of a sample of polonium. The atoms of this radioactive element disintegrate, changing into stable atoms of non-radioactive lead. The rate of this decay is proportional to the amount of polonium present. It does not depend on the age of the sample. The func-

Fig. 9.12



tion giving the amount of polonium present at any time t, therefore, has the form of Equation (6) where N_0 is the number of polonium atoms present at time zero, and k is negative.

(a) Solve the equation $\frac{1}{2} N_0 = N_0 e^{-0.005t}$ for t. The length of time required for a sample of the element polonium to decay to half its present size is called its "half life."

(b) What, approximately, is the half life of polonium that you find graphically from Fig. 9.12?

(c) How does the rate of decay when the sample is reduced in size to one-half, compare with its initial rate of decay?

6. The element uranium has a rate of decay given by

$$N' = -1.5 \times 10^{-10} N \text{ atoms per year}$$

Draw the graph of N/N_0 as a function of t, where N_0 is the initial condition.

7. Sketch the graph of $N/N_0 = e^{+0.005t}$. Is there an analogue of "half life" for exponential functions with positive exponents? (Perhaps the term would be "doubling time.")

8. To answer this question, refer to Question 7(a) in Section 9.5. A large printing press used to print cardboard posters can print just one color at a time. However, multicolored posters can be produced by running the posters through the press one time for each color. From past experience it has been determined that the percentage of rejects (blurred ink, torn paper, etc.) on a single run is never higher than 6 per cent. How many blank posters must one begin with if one needs to produce

(a) 100 one-color posters,

(b) 100 two-color posters,

(c) 100 five-color posters ?

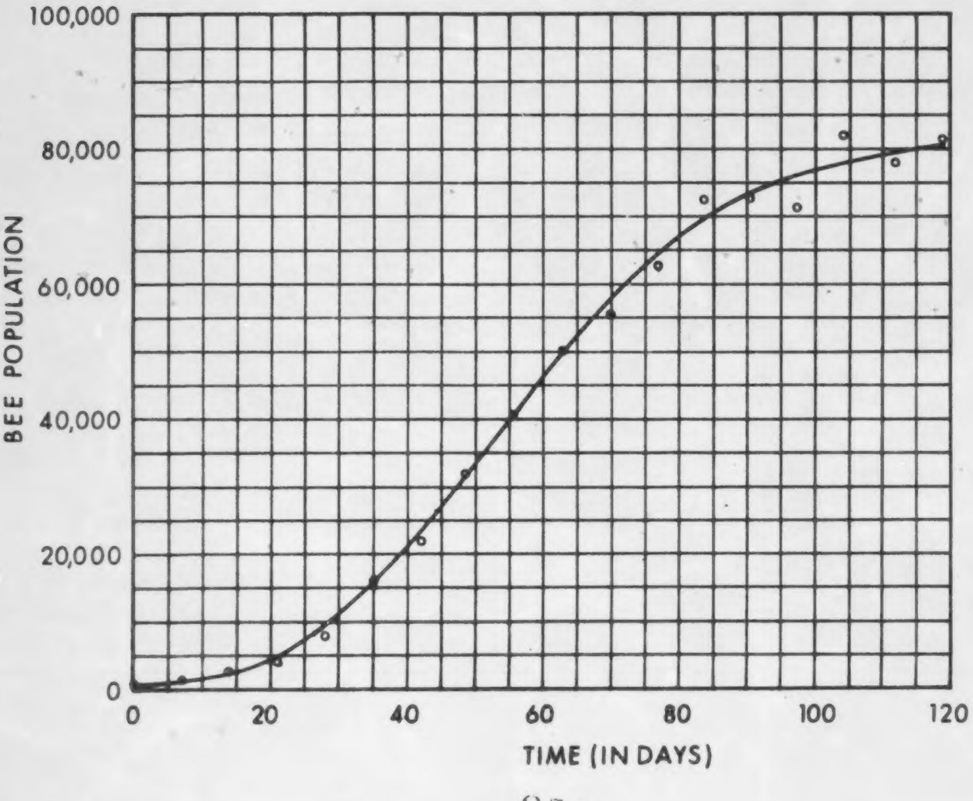
9. A population of wild rabbits (Section 9.2) may grow exponentially for some time. But clearly such exponential growth cannot continue indefinitely due to the limitations in the environment. Very often limiting factors cause populations which have appeared to grow exponentially for a while to "level off," to begin to die out, or to exhibit other erratic growth and/or decay.

Figure 9.13 shows the growth curve of a colony of bees. It is very nearly exponential for a while and then begins to level off. The growth curve of Fig. 9.13 is quite accurately described by the growth rate function

$$N' = kN\left(\frac{K - N}{K}\right)$$

where N is the number of bees at time t and k and K are constants. The factor $\left(\frac{K - N}{K}\right)$ represents the limitation on the rate of growth and

Fig. 9.13



on the ultimate size of the colony due to environmental factors. What is the significance of K?

10. (a) Find the function that represents the growth of lead formed by the decay of a sample of polonium initially containing 6×10^{23} atoms and graph the function using data from the decay curve for polonium (Fig. 9.12).

(b) Is the time it takes for the lead formed to double in amount a constant?

11. A person hears a rumor and repeats it to three other persons in one day. Assume that each of these three persons pass on the rumor to three other persons the next day who have not previously heard it. The rumor is passed on in this way for 8 days. How many persons will have heard the rumor? Is the assumption reasonable?

12. Look up the topic of C_{14} (carbon 14) dating. What is the relationship between this section and C_{14} dating?

13. (a) Use the delta process and the approximation $\ln(1 + \Delta x) \approx \Delta x$ for $x << 1$ to find $[\ln x]'$.

(b) What is the derivative of $ae^x + b \ln x$?

14. (a) Suppose $b > 0$. Find a constant k for which $b^x = e^{kx}$.

(b) What is the derivative of $f(x) = b^x$?

15. What is the derivative of $f(x) = \log x$ (the logarithm to base 10 of x)?

16. Compare the derivative of $\ln x$ with that of $\ln(cx)$. Does this comparison contradict the statement of Section 8.7 that "the antiderivatives of a given function form a family of homomorphic functions displaced vertically with respect to one another?" Why?

9.7 Recognizing Functions of the Form $y = Cb^x$

Consider the functions whose graphs are shown in Fig. 9.14(a) and 9.14(b). How can we find if they are of the form

$$y = Cb^x \tag{7}$$

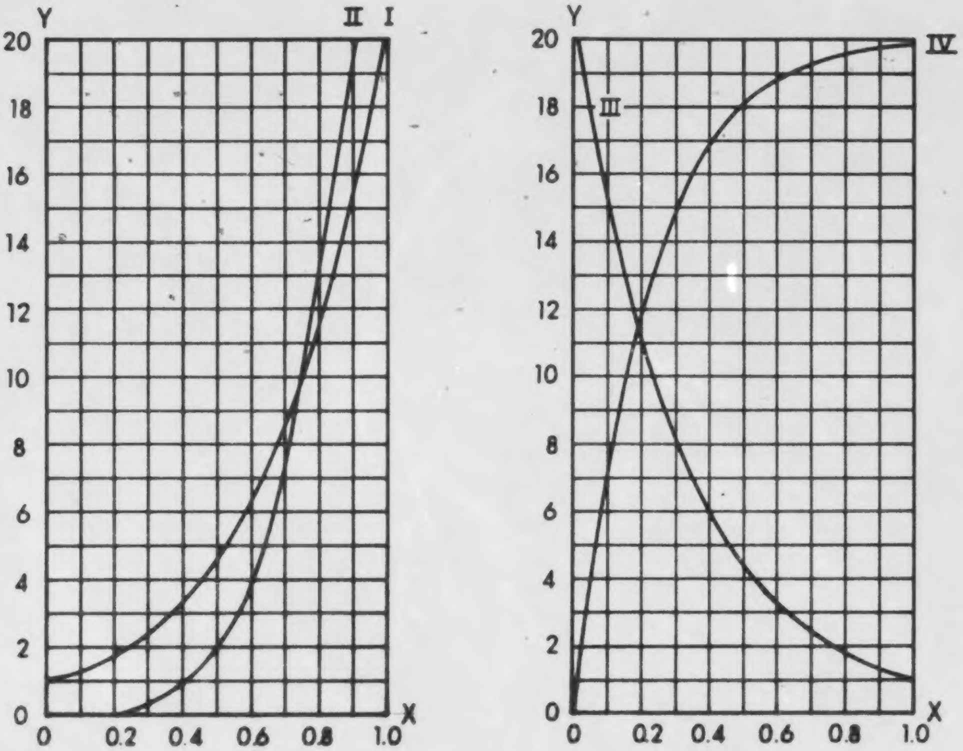
and what are the values of the constants C, and b?

If we take logarithms of both sides of Equation (7) we get the equation

$$\log y = \log C + x \log b \tag{8}$$

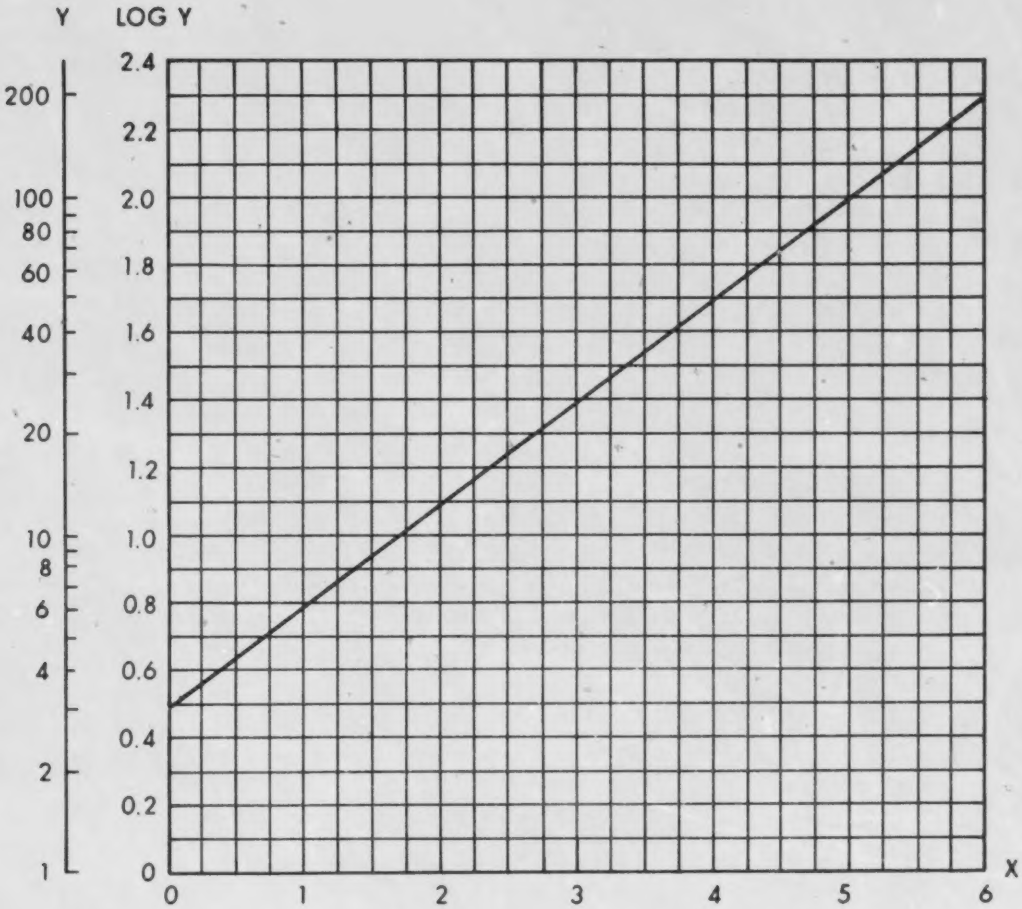
Now if we let $z = \log y$ we see that Equation (8) describes z as a linear function of x, i.e., $z = (\log b)x + \log C$. If the functions are of the form $y = Cb^x$ a plot of $z = \log y$ as a function of x will be a straight line. Figure 9.15 illustrates an example of just this situation. It is a graph of $\log y$ as a function of x for the function $y = 3(2)^x$.

Fig. 9.14



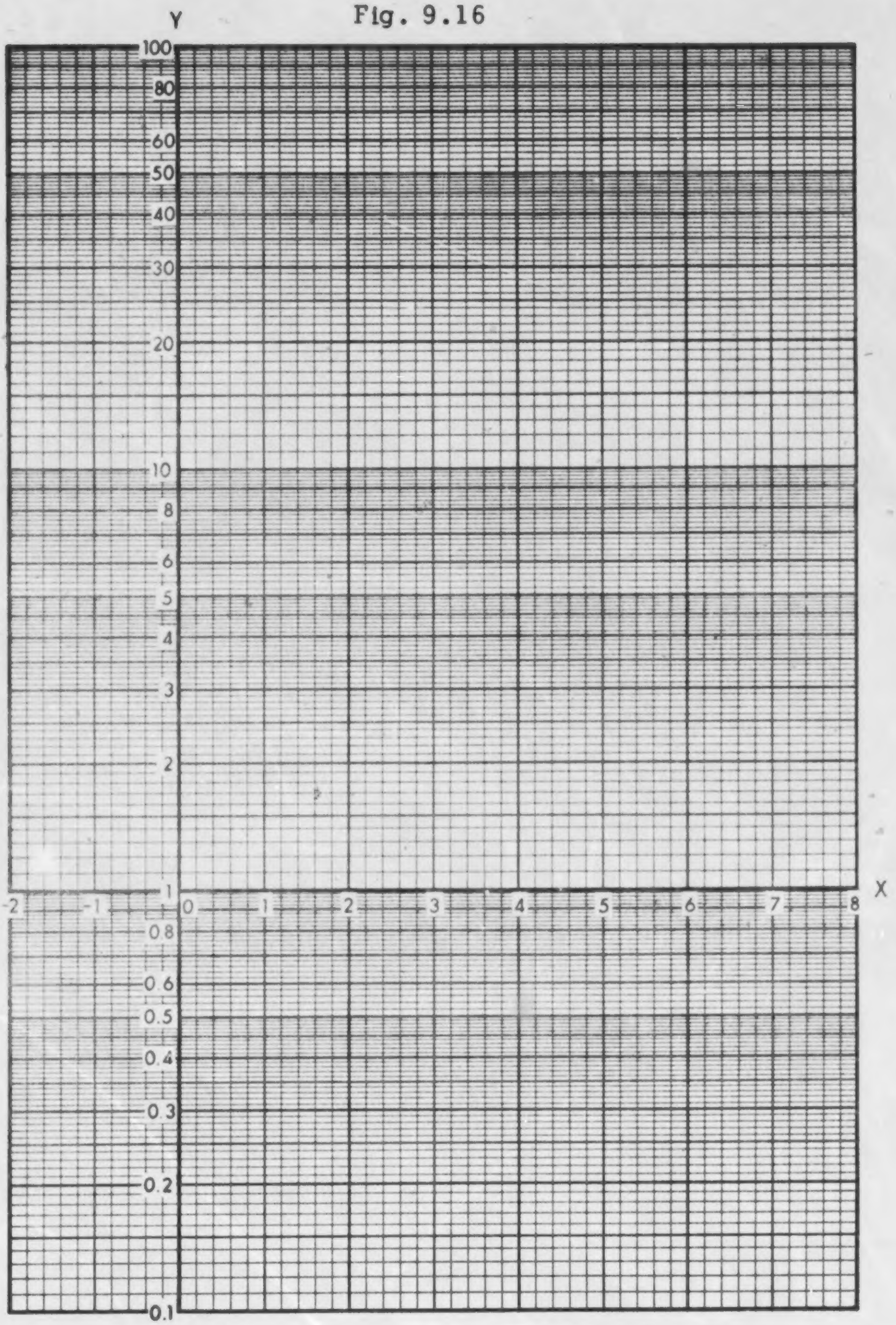
We have included on the graph of Fig. 9.15 a second vertical axis representing the numbers y whose logarithms are marked off on the log y scale. The relative displacements of the numbers on the y scale are the same as those on the C and D scale of a slide rule.

Fig. 9.15



When plotting log y versus x it is tedious to have to look up the logarithm of each value of x plotted. There is a special kind of graph paper, called semi-log paper, which eliminates this problem. A sheet of semi-log paper with x and y axes drawn on it is shown in Fig. 9.16. Note that the scale lines along the x axis are equally spaced like those on ordinary graph paper. The scale lines on the y axis are not equally spaced, however. The actual displacement from the intersection of the axes to a particular value

Fig. 9.16



on the y axis, say y_0, is proportional to the logarithm of y_0. In other words, equal displacements along the y axis are proportional to differences in the logarithms of the numbers actually marked on the scale. For example, the displacement between 1 and 10 equals the displacement between 10 and 100 since $\log 10 - \log 1 = \log 100 - \log 10 = 1$. To plot the point $(3, \log 2)$ one just goes to the 3 on the x axis and then moves up to 2 on the y axis.

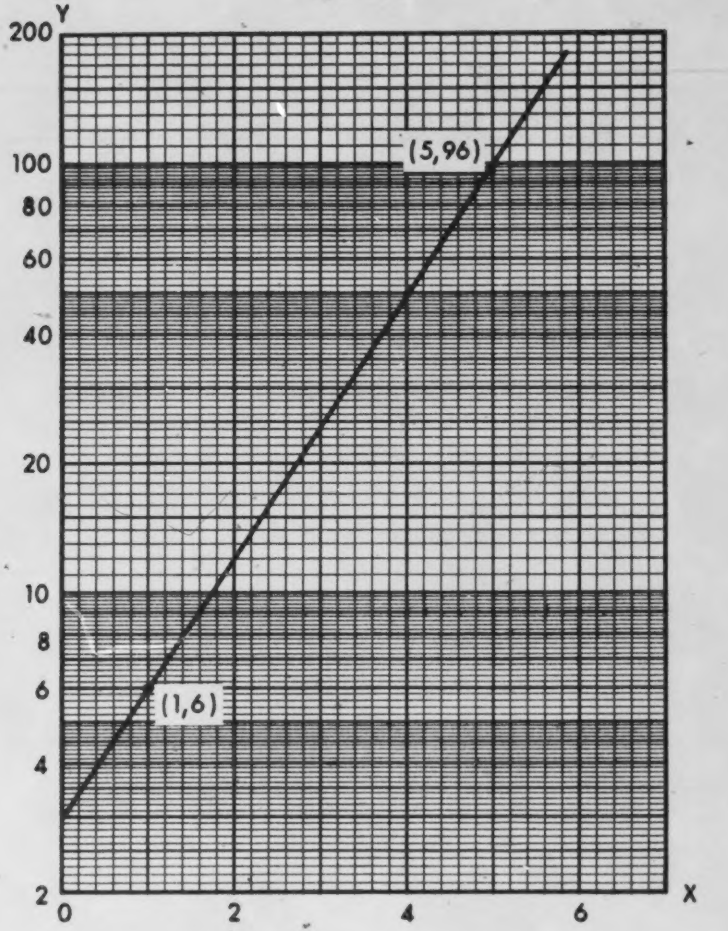
In plotting the labeled values of y versus x you are really plotting $\log y$ versus x. Semi-log paper is a convenience to help you plot $\log y$ versus x without having to use log tables just as a slide rule helps you multiply numbers by adding their logarithms without ever looking up the logarithms in a table.

To illustrate our method let us find the equation for the function in Fig. 9.15 whose graph is drawn on semi-log paper in Fig. 9.17. Since the point $(0,3)$ is on the graph we get from Equation (7) that $3 = Cb^0 = C$ so $C = 3$. Letting $x = 1$ in the equation $y = 3b^x$ gives $y = 3b$ so to find b we observe that the point on the graph with x coordinate 1 has y coordinate 6. From the equation $6 = 3b$ we have that $b = 2$. Our function therefore has as its equation $y = 3 \times 2^x$.

This method for finding C and b from the points with x coordinates 0 and 1, respectively, will always work since the line representing the graph of $y = Cb^x$ can always be extended so as to cross the vertical lines $x = 0$ and $x = 1$.

There is a limitation in using semi-log paper to plot exponential functions. One scale division on the horizontal scale can have any value you choose but the range on the vertical scale is limited. The one in Fig. 9.17, for example, can cover only a range of three consecutive decades of y values. Such paper is said to have three cycles. It can be used to plot values of y from 10^2 to 10^5 or from 10^{-4} to 10^{-1} but not from 10 to 10^5 or 10^{-3} to 10^3, etc. If you need to plot with more than three decades on the vertical axis you can attach several sheets together or you can use semi-log paper with more than three cycles.

Fig. 9.17



Questions

1. This question illustrates the fact that we have lost no generality in writing our exponential functions in this section in the form $y = Cb^x$ rather than as $y = Ce^{kx}$.

   (a) If $3^x = e^{kx}$ for every x, what is k?

   (b) If $b^x = e^{2x}$ for every x, what is b?

2. Find the equation whose graph is sketched in Fig. 9.18.

3. (a) Which of the graphs in Fig. 9.19 are graphs of a function of the form $y = Cb^x$?

   (b) For those graphs in Fig. 9.19 which represent exponential functions find the values of C and b.

Fig. 9.18

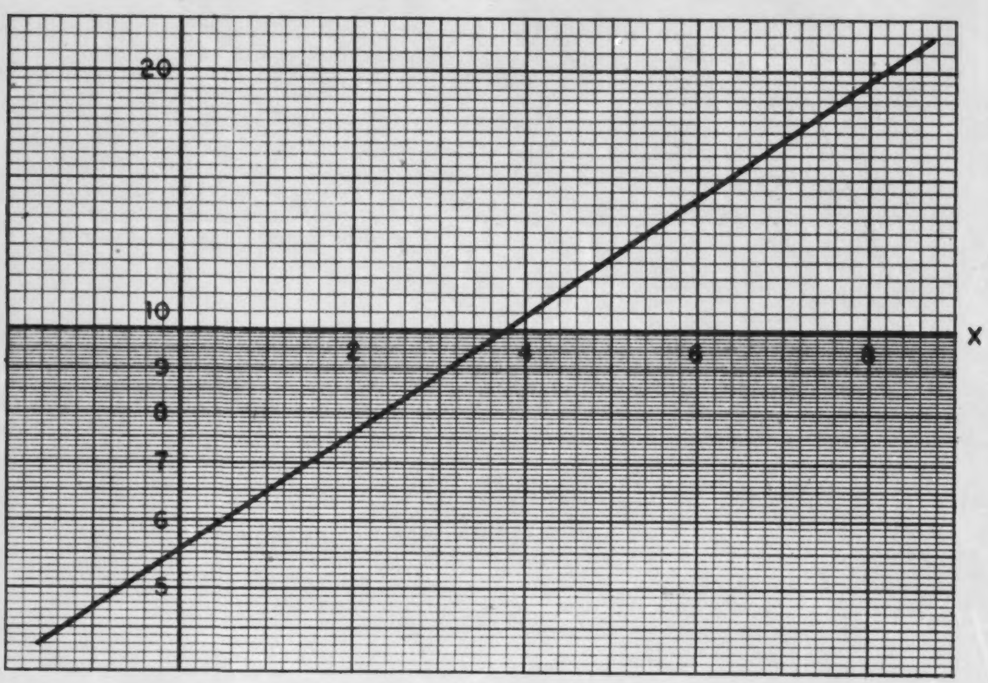
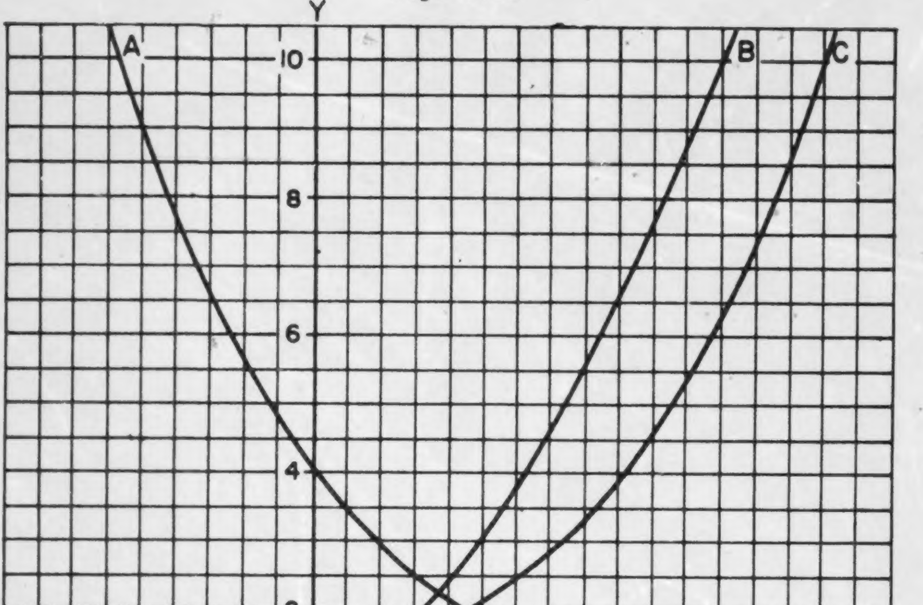


Fig. 9.19



4. The table below represents y as a function of x.

x	y
-3	2.07
-2	2.22
-1	2.67
0	4.00
1	8.00
2	20.00

(a) Plot the data on regular graph paper — the relationship should appear to be exponential.

(b) Replot the data on semi-log paper.

(c) The result of part (b) was probably disappointing. Don't give up. Find a constant d so that the set of points (x, y-d) do give a straight line when plotted on semi-log paper.

(d) Write an equation for these data.

(e) Will a function homomorphic to a function of the form $y = Cb^x$ give a straight line when plotted on semi-log paper?

5. The accompanying table gives the world population from 1650 to 1970.

Year	World Population
1650	0.545×10^9
1700	0.610×10^9
1750	0.728×10^9
1800	0.905×10^9
1850	1.17×10^9
1900	1.61×10^9
1950	2.40×10^9
1955	2.69×10^9
1960	2.92×10^9
1965	3.18×10^9
1970	3.50×10^9

(a) Make a plot of population versus year on regular graph paper.

(b) To see if the graph in part (a) is an exponential function of the form $y = Cb^x$, plot log y versus x on semi-log paper.

(c) Notice in the graph of part (b) that the portion from 1950 to 1970 is fairly linear. Replot this portion with a larger scale along the x axis.

(d) From your graph in part (c), find the values of b and C, and write the exponential function that describes the population growth from 1950 to the present.

(e) Demographers project a world population of 6.27×10^9 by the year 2000. Extrapolate your graph in part (c) to the year 2000, and compare your result with this figure.

6. Make a semi-log graph of the growth curve of the bee population shown in Fig. 9.13.

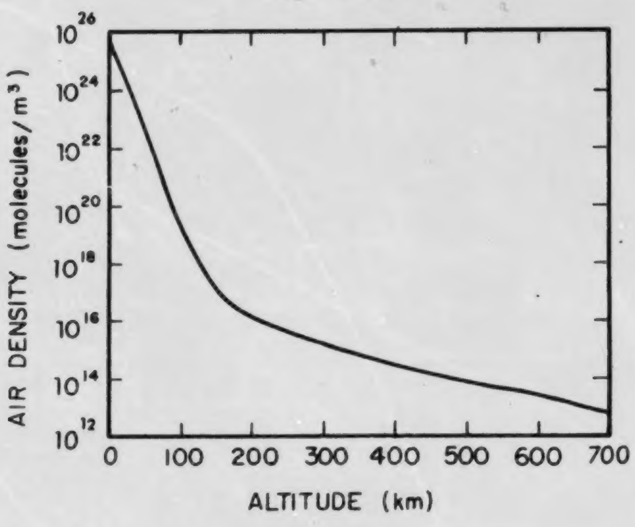
(a) For about how many days is the growth of the colony exponential?

(b) During the exponential growth of the colony what is the time interval during which the bee population doubles in size?

(c) Use the equation given in Question 8, Section 9.6, to find an equation for the maximum size of the colony of bees described in the question.

7. Figure 9.20 is a graph of the density of the atmosphere as a function of altitude. Here the density is displayed over a range of nearly seven orders of magnitude. Is the function exponential?

Fig. 9.20



8. In the spring of 1937, eight ring-neck pheasants were introduced on a protected island off the coast of the state of Washington. Each spring a count of their population was made. The results are shown in the table below. Did the colony grow exponentially?

Year	Population
1937	8
1938	30
1939	90
1940	300

9. Which of the curves in Fig. 9.14(a) and (b) are of the form $y = Cb^x$? For those that are, write their equations in the form $y = e^{kx}$.

10. Replot Fig. 6.2 using semi-logarithmic paper. What do you conclude?

9.8 Recognizing Functions of the Form $y = m\log x + b$

In the preceding section we used semi-log paper to identify exponential functions. We can also use semi-log paper to identify and specify logarithmic functions. If we want to determine if a function has the form

$$y = m\log x + b \tag{8}$$

we plot y as a function of $z = \log x$. If the function indeed has the form of Equation (8) we will again obtain a straight line.

We can use semi-log paper to plot Equation (8) with the logarithmic scale for the x axis as in Fig. 9.21. For example, this time the point $(\log 5, 3)$ is plotted as simply $(5,3)$ on the semi-log paper. From inspection we see that the y-intercept in Fig. 9.21 is 2.18. (Note that this is the value of y when $x = 1$.)

The slope of the straight line in Fig. 9.21 is $m = \frac{\Delta y}{\Delta \log x}$. By choosing points for which $\log x$ is easy to compute we can find m without using a log table. For example, since it appears that $(1, 2.18)$ and $(10, 3.35)$ are on the graph we have that

$$m = \frac{\Delta y}{\Delta \log x} = \frac{3.35 - 2.18}{\log 10 - \log 1} = 1.17$$

so the equation whose graph is given by Fig. 9.21 is $y = 1.17 \log x + 2.18$.

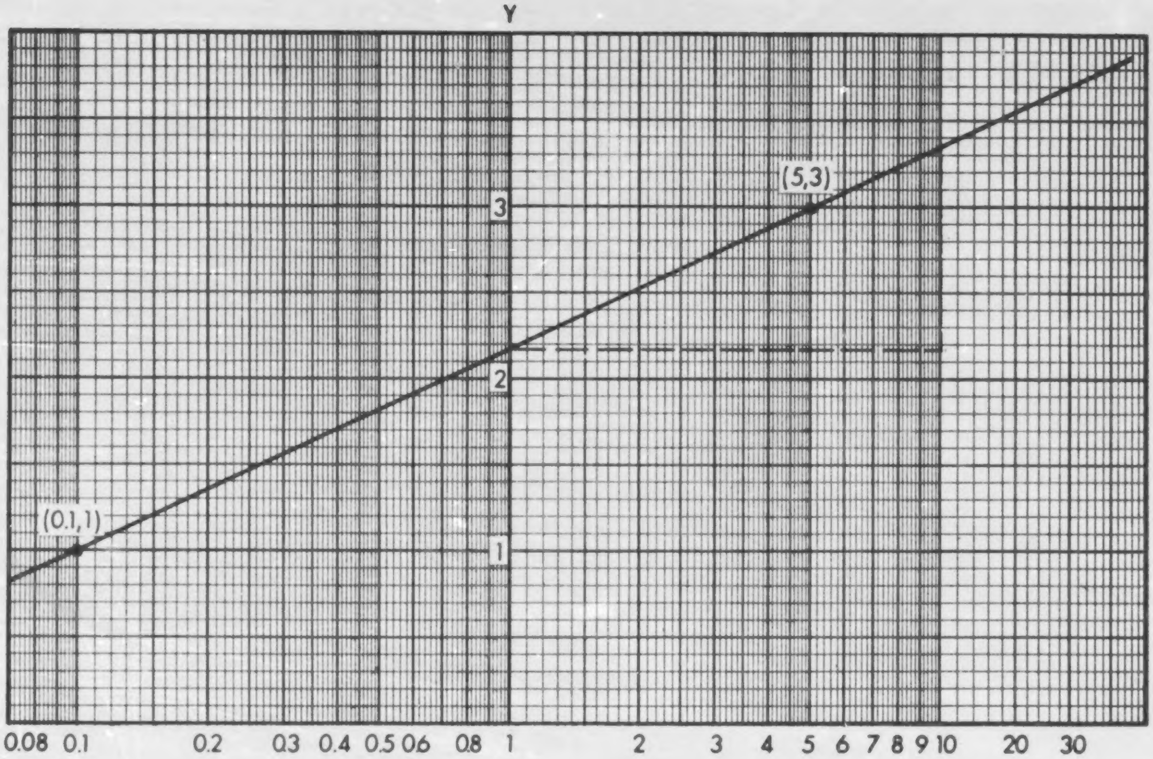


Fig. 9.21

Questions

1. The following table represents y as a function of x.

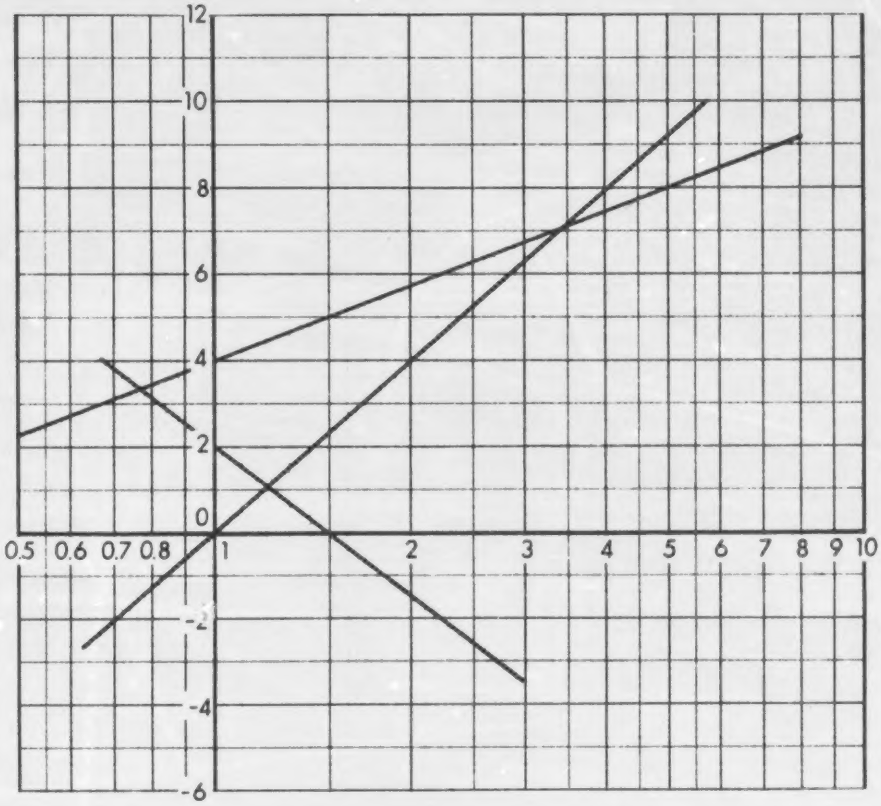
x	y
1	4.50
4	6.10
10	6.95
15	7.40
30	8.15
50	8.45

(a) Find the "best" functions of the form $y = m\log x + b$ to fit this data.

(b) Estimate the error involved in using your function to predict the value of y corresponding to a particular value of x rather than consulting the table.

2. Find the equations for the functions whose graphs are plotted on semi-log paper in Fig. 9.22.

Fig. 9.22



3. Plot y versus $\log x$ for Tables I and II. Decide in each case whether the listed data can possibly correspond to a function of the type $y = m\log x$. If so, find the value of m.

TABLE I

x	y
0.02	7.30
0.08	4.72
0.40	1.71
1.00	0.00
6.00	3.35
20.00	5.59
60.00	7.65
100.00	8.60

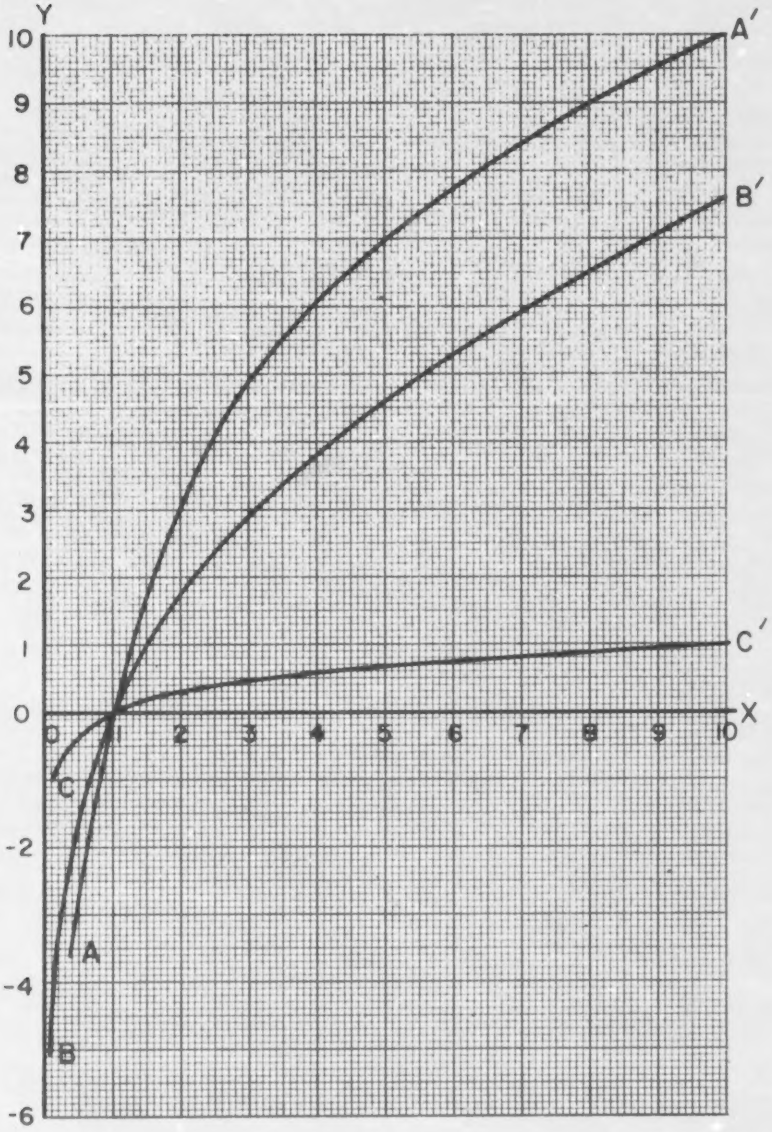
TABLE II

x	y
0.02	-3.30
0.08	-2.02
0.40	-0.67
1.00	0.00
6.00	1.13
20.00	1.75
60.00	2.30
100.00	2.52

4. (a) Which of the graphs in Fig. 9.23 are graphs of logarithmic functions of the form y = mlog x + b?

(b) For those graphs in Fig. 9.23 which were identified as graphs of logarithmic functions in part (a) find the values of m and b.

Fig. 9.23



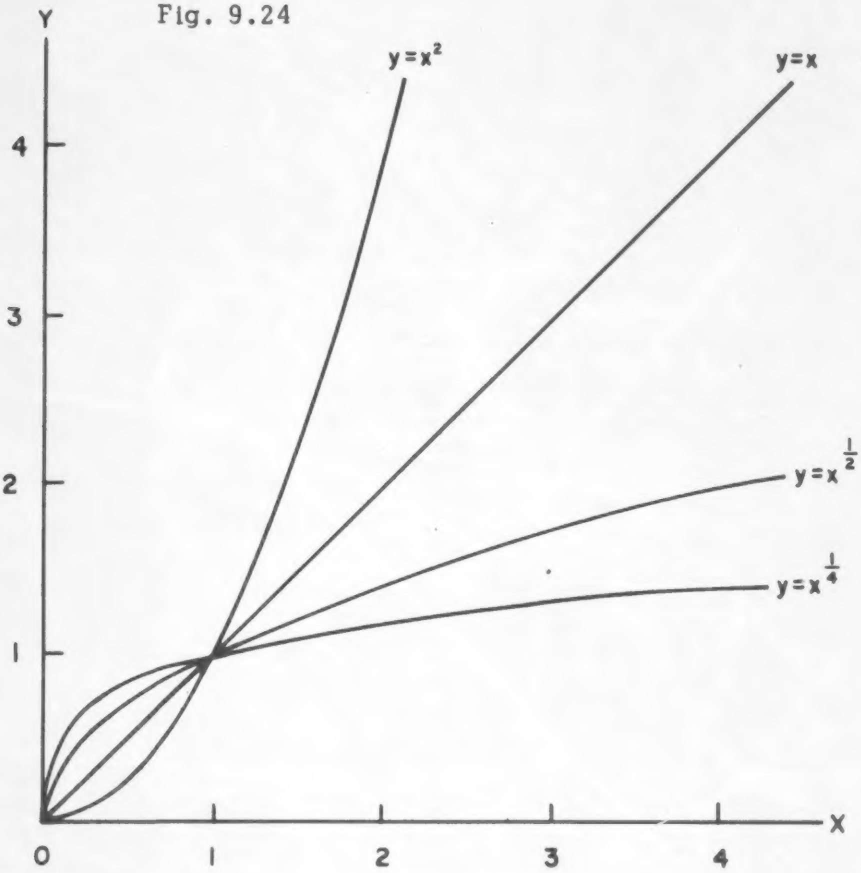
9.9 Recognizing Functions of the Form $y = ax^n$

In Chapter 7 we considered functions of the type $y = ax^n$, where n was an integral number. We can also consider functions of the form

$$y = ax^n$$

where x is greater than zero and n is any number, integral or non-integral. Figure 9.24 shows the shape of several graphs corresponding to $a = 1$ and different values of n.

Fig. 9.24



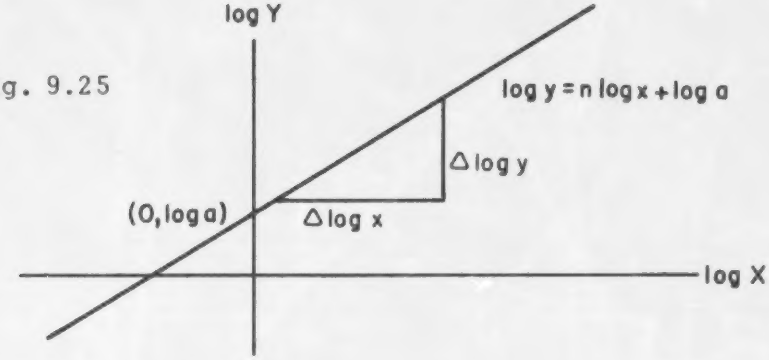
We have already described in Chapter 7 how to investigate whether a table of values of x and y represents a function $y = ax^n$, where n is some integer. For example, if we suspect a relation of the form $y = ax^2$, we plot y versus the quantity x^2. With the aid of logarithms we can now apply a more general method which will enable us to decide whether the values of x and y given in a table of data describe any function of the type $y = ax^n$ and, if so, what the values of a and n happen to be.

We shall first assume that a is positive. If the values in a table of data satisfy $y = ax^n$, as we have just defined it, all the given values of y and x are positive and we can have a relation between $\log x$ and $\log y$. Taking the logarithm of both sides of $y = ax^n$, we find that

$$\log y = n \log x + \log a$$

Therefore, if we plot $\log y$ as a function of $\log x$, we will get a straight line with slope $n = \frac{\Delta(\log y)}{\Delta(\log x)}$ and a vertical intercept $\log a$ (Fig. 9.25).

Fig. 9.25



To avoid using a table of logarithms we can plot the values of x and y on "log-log" graph paper, which differs from semi-logarithmic paper by having a logarithmic scale along both the x and y axes.

Let us investigate the nature of the function represented by the data in Table 9.3.

TABLE 9.3

x	y
0	0
0.5	0.15
1.0	0.48
1.5	1.00
2.0	1.63
3.0	3.30
4.0	5.30
5.0	8.00
6.0	10.90
7.0	14.40

Figure 9.26 is a graph of y as a function of x for this table. From the shape of the graph it is plausible that the corresponding function is of the type $y = ax^n$, so we plot y as a function of x using log-log paper (Fig. 9.27), numbering the scales in the same way as described in Sections 9.7 and 9.8, where the use of semi-logarithmic paper was discussed. The x axis and the y axis cross at the point we have labeled (1, 1), corresponding to (log 1, log 1) = (0, 0). If we wish to plot the point (1.5, 1.00) from Table 9.3, we find the intersection of the vertical line numbered 1.5 and the horizontal line marked 1.00. Note that the point (0, 0) cannot be plotted. Plotting the remaining points, we see that the graph is a straight line, so we know that a function of the form

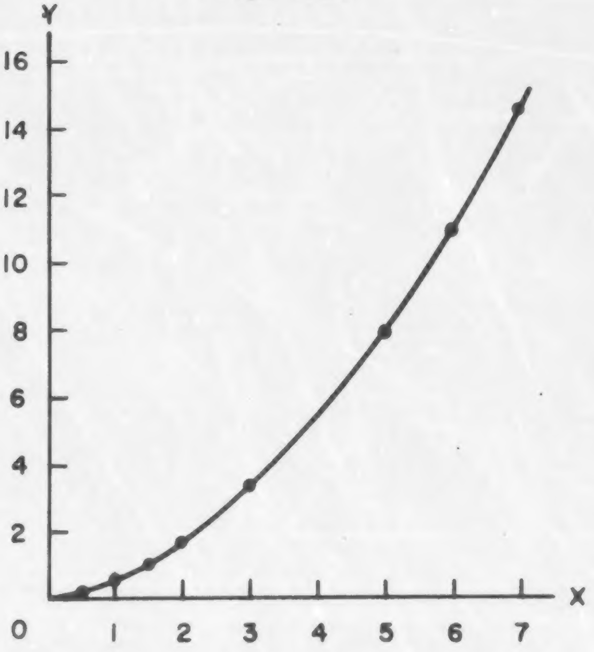
$$\log y = n \log x + \log a$$

describes the data in Table 9.3. The value of the slope is

$$n = \frac{\Delta(\log y)}{\Delta(\log x)}$$

On log-log graph paper, displacements in inches, centimeters, etc., on the paper are proportional to the corresponding differences in logarithms

Fig. 9.26



of those numbers marked on the scale. Therefore, in Fig. 9.27

$$n = \frac{\Delta(\log y)}{\Delta(\log x)} = \frac{k\overline{BC}}{k\overline{AC}} = \frac{\overline{BC}}{\overline{AC}}$$

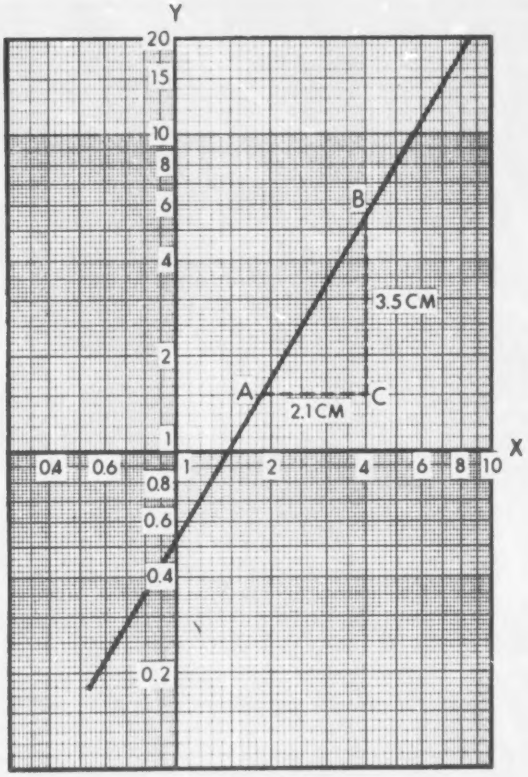
where k is the constant of proportionality between logarithms and displacements. This means that we can find the slope on log-log graph paper by taking the ratio of the actual displacements Δy and Δx measured in centimeters on the paper; we do not have to find the logarithms of any numbers. Measuring BC and AC in Fig. 9.25 gives

$$n = \frac{BC}{AC} = \frac{3.5\text{ cm}}{2.1\text{ cm}} = 1.7$$

To find a in the function $y = ax^{1.7}$, we look at the y intercept in Fig. 9.27. It shows that when $x = 1$, $y = 0.52 = a(1^{1.7}) = a$. Therefore, we conclude that the numbers in the table satisfy the function

$$y = 0.52x^{1.7}$$

Fig. 9.27



To help you become familiar with log-log paper and the techniques outlined above, we have plotted the functions shown in Fig. 9.24 on log-log paper (Fig. 9.28). Because the original functions are of the form $y = ax^n$ we know that a log x versus log y plot will give a straight line, as Fig. 9.28 indeed shows.

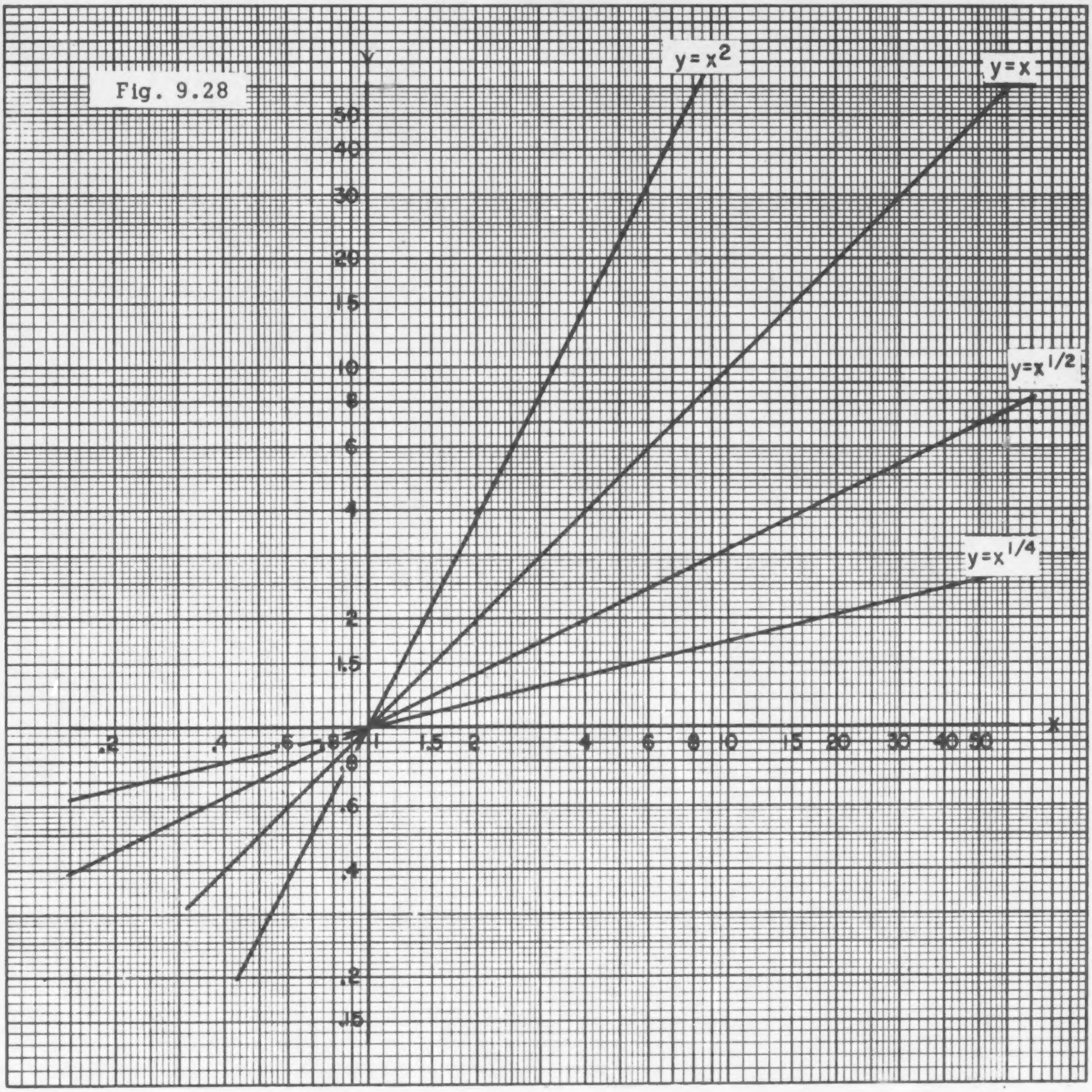


Fig. 9.28

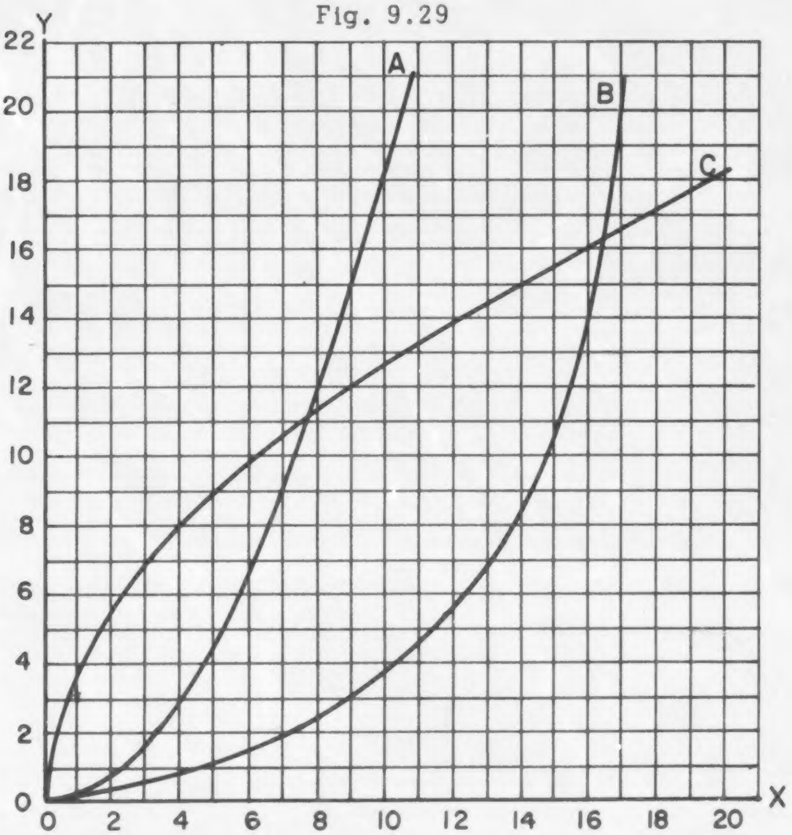
So far we have dealt only with the case in which a is positive and both x and y are positive. This is necessary because we can only plot points on log-log paper that fall in the first quadrant of an ordinary graph. If a is negative and we make the restriction that x is always positive, then y must be negative. This corresponds to a graph lying entirely within the fourth quadrant of an ordinary graph. We cannot plot such a function on log-log paper, but instead we can plot the function $y = -(ax^n)$.

Questions

1. Determine the slopes of the straight lines shown in Fig. 9.28 to convince yourself that they agree with the value of n given for each function.
2. (a) Use log-log paper to ascertain if the curves in Fig. 9.29 are of the form $y = ax^n$.

   (b) For the curves that are of the form $y = ax^n$, find a and n.

Fig. 9.29



3. Plot the function $y = ax$ on log-log graph paper for several values of a.

4. You know that a straight line on a sheet of log-log paper corresponds to a function of the general form $y = ax^n$. How does each of the following conditions restrict the values of a and n? The graph
   - (a) has negative slope,
   - (b) has slope zero,
   - (c) has a y intercept greater than 1,
   - (d) passes through the "origin" (the point (1, 1)).

5. Can one always find the slope of a line on log-log paper by measuring the vertical and horizontal displacements with a ruler and finding the ratio of the two? (Does it matter what units the two displacements are measured in? Would it matter if the graph paper had a different displacement for one cycle along the x axis than for one cycle along the y axis?)

6. The following table, the result of an experiment, gives values for the force of repulsion F between two electrically charged spheres as a function of the distance d between their centers:

Distance, d (cm)	Force, F (arbitrary units)
3.4	7.3
3.8	6.1
4.4	5.2
4.7	4.0
5.4	3.4
6.2	2.7
6.9	2.0
7.9	1.7
8.7	1.2
10.6	0.7
13.1	0.5

(a) Make a graph from the values in the table using log-log graph paper.

(b) Both d and F were measured to ±0.05 units. Draw error rectangles around each data point. Why are the rectangles not of the same size?

(c) Is the relation between F and d of the form $F = kd^n$, where k is a constant? (Note that both d and F are given to only two significant digits.)

(d) Compare your value of n with those of your classmates by making a class histogram. What is the best class value of n?

9.10 Scale Stretching by Logarithmic Plotting

In each of the preceding three sections we have made use of logarithmic plotting, i.e., we have chosen to plot the logarithms of at least one of the variables rather than the actual values of the variable. In each case we were able to use some kind of logarithmic plotting to determine the form of a certain kind of function from its graph.

Another important use of logarithmic plotting arises from the fact that a graph of y versus log x, rather than versus x, can be used to stretch out the portion of the x axis corresponding to small values of x. This is useful in some cases for clarity of display even when no correspondence with any logarithmic function is suspected. (The y axis can similarly be stretched out for the smaller values of y by this same method.) This means that we can plot data ranging over several powers of ten with the axis scales expanded for the smaller powers of ten. For example, suppose we wish to plot the curve passing through the points given in Table 9.4.

TABLE 9.4

y	x
0	0.06
1.0	0.13
2.0	0.53
2.5	1.00
3.0	2.60
3.5	10.00
3.8	32.00
4.0	100.00

Due to the large range of values for x, if we plot y versus x we must use such a large value of x per scale division that the lower part of the curve is nearly indistinguishable from the y axis (Fig. 9.30).

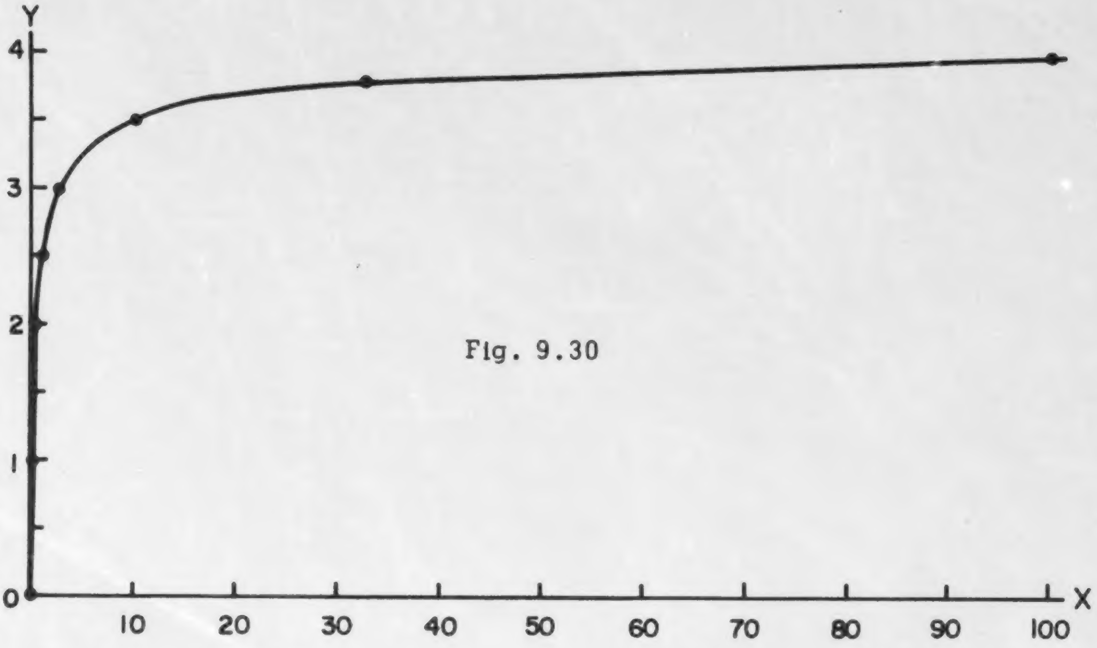


Fig. 9.30

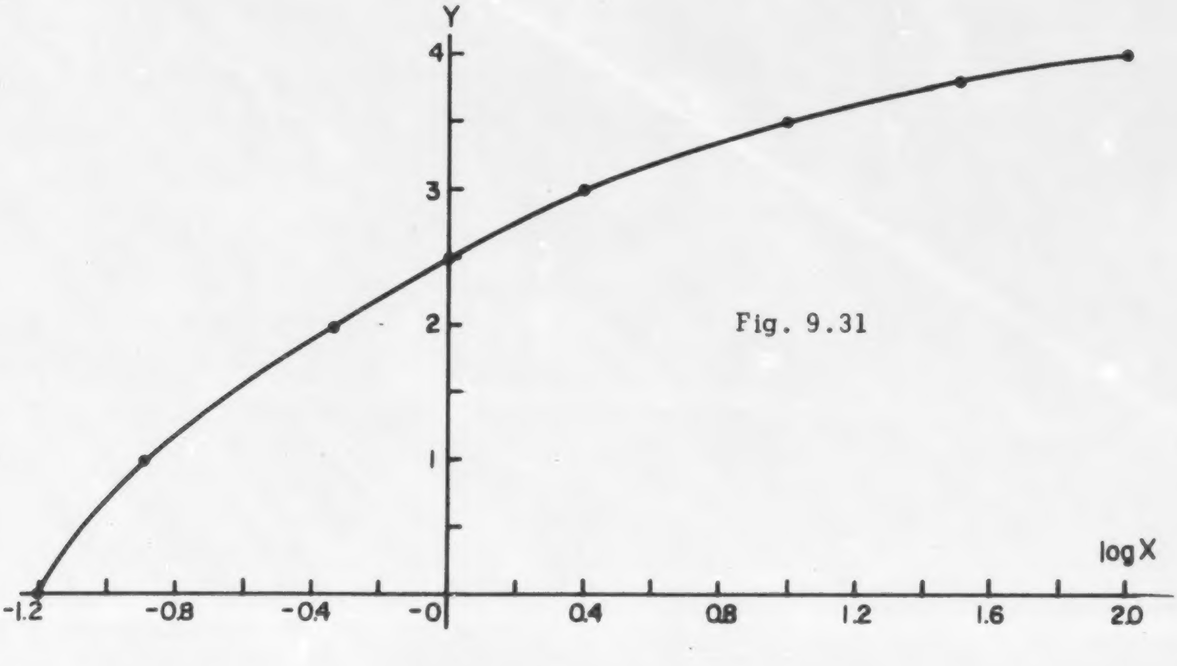


Fig. 9.31

If we plot y versus log x, however, the horizontal axis need go only from log (0.06) = -1.2 to log (100.00) = 2 and the data points are much more evenly spaced (Fig. 9.31). A quick way of making such plots is to plot y versus x on semi-log paper using the logarithmic scale for the x axis.

Questions

1. In the table below T is the time in years it takes the planet to make one orbit around the sun and R is the distance in kilometers from the planet to the sun. Use the table to plot

(a) $\frac{R}{T}$ as a function of T,

(b) $\frac{R}{T}$ as a function of T on semi-log graph paper with the logarithmic scale on the T-axis.

Planet	T	$\frac{R}{T}$
Mercury	0.24	24.0×10^7
Venus	0.61	18.0×10^7
Earth	1.00	15.0×10^7
Mars	1.90	12.0×10^7
Jupiter	12.00	6.6×10^7
Saturn	29.00	4.9×10^7
Uranus	84.00	3.4×10^7
Neptune	165.00	2.7×10^7
Pluto	248.00	2.4×10^7

2. (a) Plot the graph of Fig. 6.12 on semi-log graph paper, using the logarithmic scale for the x axis and the three decades from 1 to 1000.

(b) What is gained by a semi-log graph compared to the original graph?

3. The distance that electrons can penetrate through a substance depends on the substance and the energy of the electrons (which depends on their speed). The table on page 266 gives the range-energy relation for the penetration of electrons into aluminum.

Energy, E (Mev)	Range, R (cm)
4.2×10^{-2}	1.0×10^{-3}
8.5×10^{-2}	3.7×10^{-3}
1.0×10^{-1}	4.8×10^{-3}
2.0×10^{-1}	1.6×10^{-2}
4.0×10^{-1}	4.8×10^{-2}
1.0	1.5×10^{-1}
2.0	3.4×10^{-1}
3.0	5.4×10^{-1}
4.0	7.4×10^{-1}
5.0	9.5×10^{-1}

Plot both $\underline{E}$ as a function of $\underline{R}$ and, on log-log graph paper, log E as a function of log R. Which graph gives the best display of the data in the table?

Chapter 10. THE SINE AND COSINE FUNCTIONS

In Fig. 10.1, if the angle θ remains the same but we choose different values of the hypotenuse r, we have a family of similar right triangles. In these triangles, the ratios of corresponding sides are equal.

If, on the other hand, we draw a family of right triangles with the same base x_1, as in Fig. 10.2, these triangles are not similar.

Here, the ratio $\frac{y_1}{r_1}$ depends on the value of θ_1. The ratio $\frac{y_2}{r_2}$ depends on the value of θ_2, etc.

Fig. 10.1

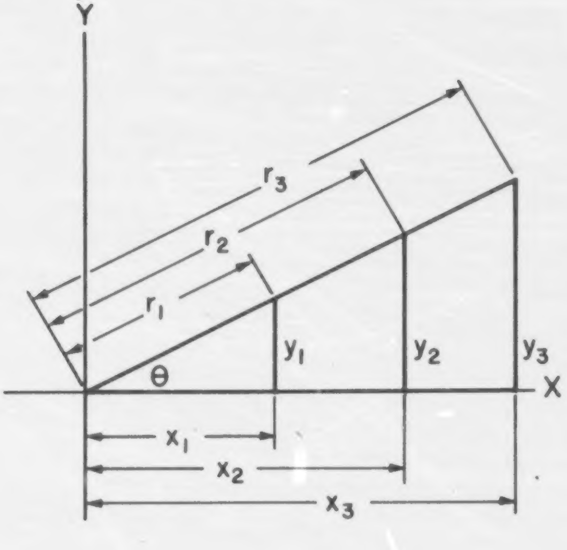
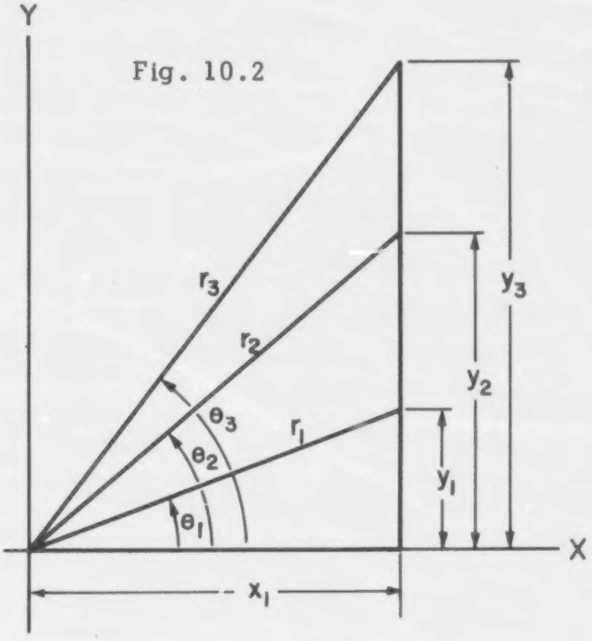


Fig. 10.2



From Figs. 10.1 and 10.2 we see that there is a clear relationship between the size of an angle and the ratio of certain sides of the right triangle that contains the angle. This chapter deals with two such relationships, the sine function and the cosine function. We begin our study with some observations about angles.

10.1 Sectors and Radians

All circles are similar, but when are two sectors of circles similar? It is evident from Fig. 10.3 that two sectors are similar when their central angles (θ in Fig. 10.3) are equal. In similar figures the ratios between corresponding parts are equal. For similar sectors, in particular, the ratio of the lengths of the arcs equals the ratio of the corresponding radii.

In Fig. 10.3

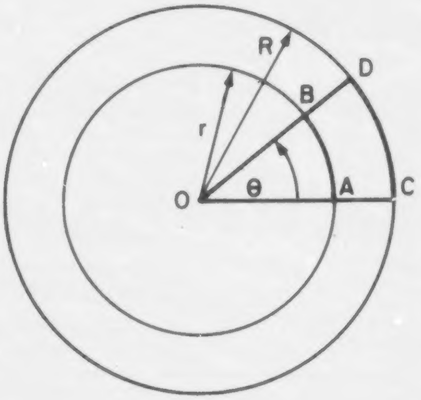
$$\frac{CD}{AB} = \frac{R}{r}$$

so

$$\frac{CD}{R} = \frac{AB}{r}$$

Thus, in similar sectors the ratio of the arc length to the radius is constant; it is independent of the radius.

Fig. 10.3



This suggests that the ratio of arc to radius is a convenient measure for the central angle. The unit of measuring angles in this way is called a radian;

$$\theta \text{ (in radians)} = \frac{\text{arc}}{\text{radius}}$$

Since an angle is a ratio of two lengths, it is independent of the unit of length used. It is a pure number.

Figure 10.4 shows an angle equal to 1 radian and one equal to 0.1 radian. Since the circumference of a circle of radius r is $2\pi r$, a full turn or 360° equals $\frac{2\pi r}{r} = 2\pi$ radians.

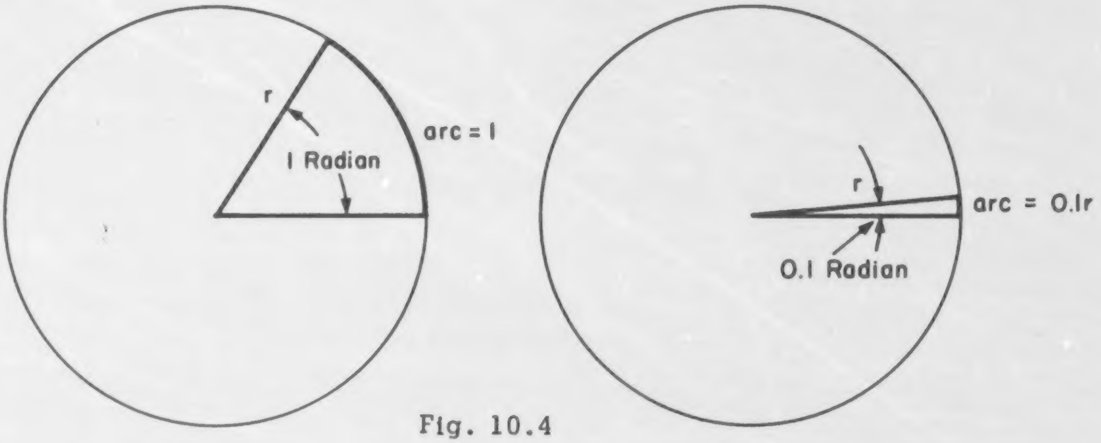


Fig. 10.4

An angle expressed in radians is often written by omitting the unit. Thus, an angle given as a number only is always understood to be in radians. For example, an angle of π radians is usually said to be of size π. An angle of 2 radians is written as just 2.

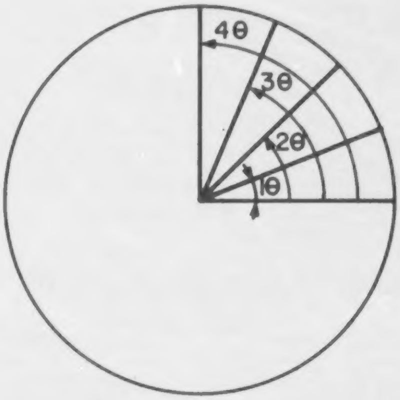


Fig. 10.5

Measuring the central angle in radians provides a simple formula for the area of a sector. As seen from Fig. 10.5, the area of a sector is proportional to the central angle. The area of a sector of central angle 1 radian is $\frac{1}{2\pi}$ of the area of the circle or $\frac{1}{2\pi} \cdot \pi r^2 = \frac{1}{2} r^2$. Therefore, the area of a sector of central angle θ is

$$A = \frac{1}{2}\theta r^2$$

Questions

1. How many degrees equal 1 radian?
2. Angles of 30°, 45°, 60°, 90°, and 180° occur frequently. Express them in radians.
3. What is the formula for the area of a sector when the central angle is given in degrees?
4. A right triangle has one leg equal to the radius of a circle and the other leg equal to the circumference of a circle.
   (a) What is the ratio of the lengths of the two legs?
   (b) How does the area of this triangle compare with the area of the circle?
5. Find the area of the shaded portion of the figure in Fig. 10.6.

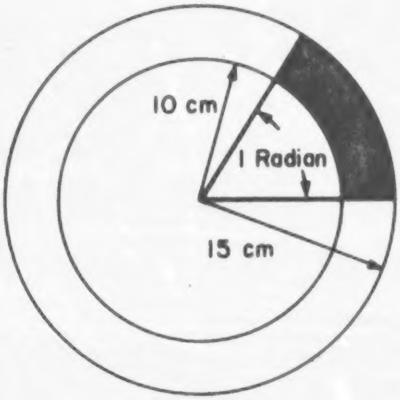


Fig. 10.6

6. The length of the chord subtended by a small central angle is approximately equal to the length of the arc it subtends. Also, the smaller the angle, the better the approximation. You can test these statements by the following procedure: Draw a semicircle of large radius. Using your value of R, make a table of

Angle	Arc Length	Length of Chord	Fractional Difference

for different angles by successively bisecting the central angle about 5 times. In each case the arc length can be calculated and the length of chord can be measured with a ruler.

7. The thumb and the outstretched hand are useful instruments for approximate angle measurements. What angle does the width of your thumb subtend when you stretch your arm out?
8. The moon subtends very nearly the same angle from the earth's surface as does the sun. (Think of a total solar eclipse.) The moon is about 2.5×10^5 miles away, and the sun is about 10^8 miles away; what is the ratio of their diameters?
9. The moon is 2.5×10^5 miles away and subtends an angle of 0.01 radians from the earth. If it were 4×10^7 miles away, how large an angle would it subtend?

10.2 Definitions

We say an angle is in standard position if its vertex is at the origin O of the coordinate system and its initial side $\overline{OA}$ extends along the positive x axis (Fig. 10.7). If, in Fig. 10.7, (x,y) are the coordinates of Q, the point of intersection of the terminal side $\overline{OB}$ of the angle AOB and the circle,

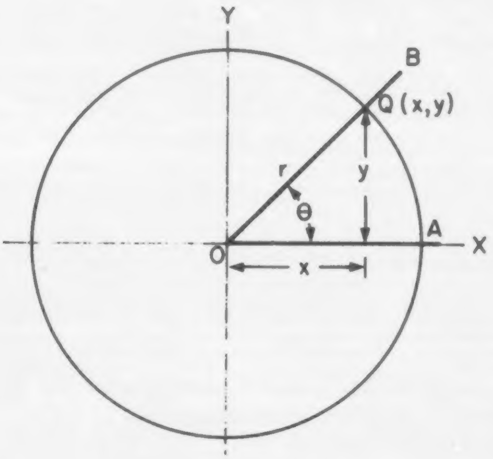


Fig. 10.7

we define the functions sine θ and cosine θ such that

$$\sin \theta = \frac{y}{r}$$

$$\cos \theta = \frac{x}{r}$$

where sine and cosine are abbreviated to sin and cos respectively. Notice that x and y are positive or negative depending on which quadrant in the coordinate system each is located. They are, for example, both positive as depicted in Fig. 10.7 but, depending on the size of θ, they can be negative. The radius r, however, is always taken to be positive.

Referring to the figure again, we see from the geometry of the diagram that

$$r^2 = x^2 + y^2$$

so

$$r = \sqrt{x^2 + y^2}$$

Thus we can write

$$\sin\theta = \frac{y}{\sqrt{x^2 + y^2}}$$

and

$$\cos\theta = \frac{x}{\sqrt{x^2 + y^2}}$$

Since sin θ and cos θ are both functions of θ, there must be a way to express one in terms of the other. Indeed, from the last two equations it follows that:

$$\sin^2\theta + \cos^2\theta = \frac{y^2}{x^2 + y^2} + \frac{x^2}{x^2 + y^2} = \frac{x^2 + y^2}{x^2 + y^2}\,*$$

or

$$\sin^2\theta + \cos^2\theta = 1$$

From this it follows that

$$\sin\theta = \sqrt{1 - \cos^2\theta} \quad \text{and} \quad \cos\theta = \sqrt{1 - \sin^2\theta}$$

We can, by constructing graphs of the sine and cosine, find their values for all angles between 0 and 2π. To construct these graphs we proceed as follows:

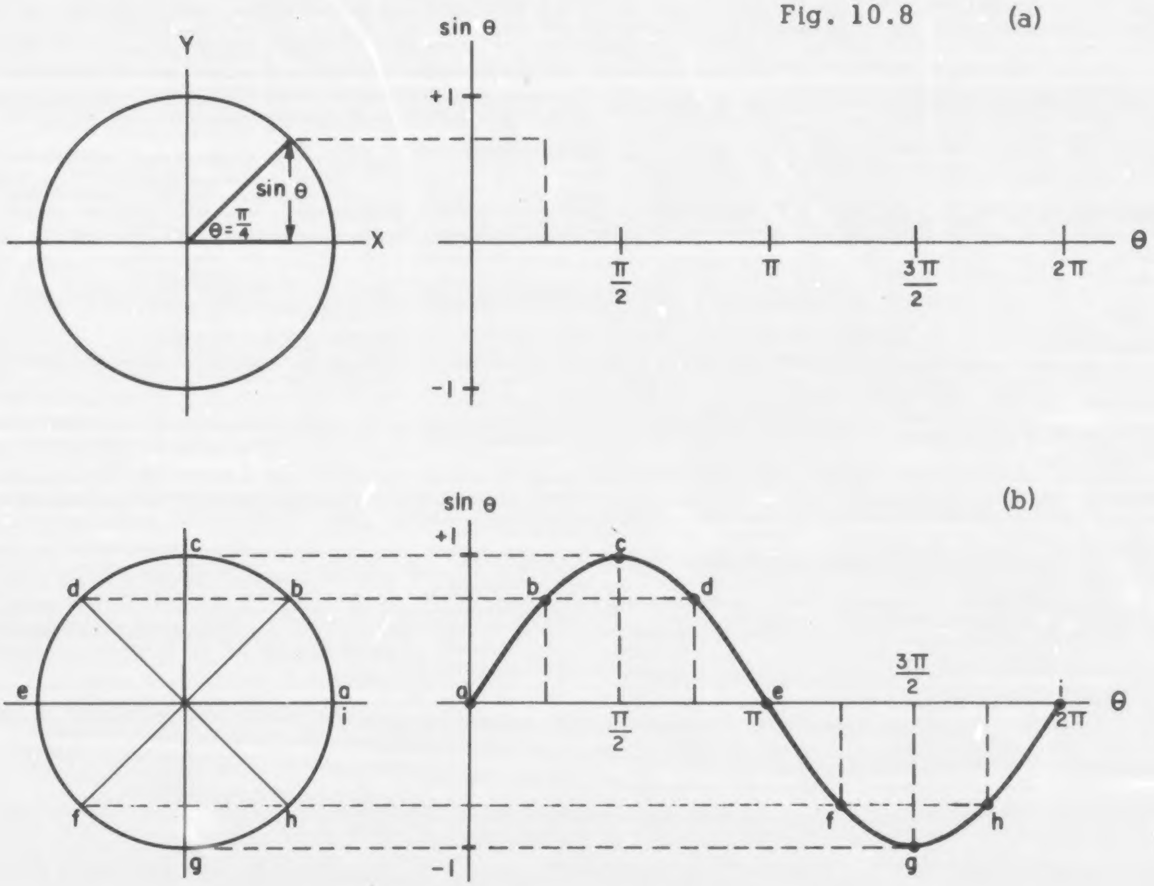
Using a ruler and a compass, we construct what is called a unit circle (Fig. 10.8(a)). That is, regardless of the actual length of the radius, we label it 1 and call this length 1 unit. If the angle $\frac{\pi}{4}$ is drawn by bisecting the first quadrant with a compass, then the coordinates of the point it intersects on the circle are equal to its sine and cosine. This follows from the observation that for r = 1, $\sin\theta = \frac{y}{r} = \frac{y}{1} = y$ and $\cos\theta = \frac{x}{r} = \frac{x}{1} = x$, so y and x have the same numerical values as $\sin\theta$ and $\cos\theta$.

---

*The notation $\sin^2\theta$ means take the sine of θ and square it: that is, $\sin^2\theta = (\sin\theta)^2$. Whereas, $\sin\theta^2$ means square θ and then take the sine of the result.

If we construct scales as shown in Fig. 10.8(a), we can mark off points directly from the unit circle. Figure 10.8(a) shows this process for $\theta = \frac{\pi}{4}$, where we see that $\sin\frac{\pi}{4} \approx 0.7$. In Fig. 10.8(b) we have constructed other angles and have marked off their sines on the scale. When sufficient points have been located, they are connected by a smooth curve. This gives the graph for $\sin\theta$ as a function of the angle θ as shown at the right in Fig. 10.8(b).

Fig. 10.8

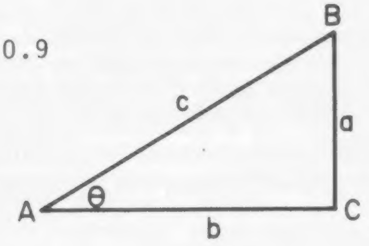


Questions

1. If $0 < \theta_1 < \theta_2 < \frac{\pi}{2}$, which is larger:
   (a) $\sin \theta_1$ or $\sin \theta_2$?
   (b) $\cos \theta_1$ or $\cos \theta_2$?
2. Plot the following points on a coordinate system, and find for each point the value of r, $\sin \theta$, and $\cos \theta$, where θ is in standard position.
   (a) (3,4)
   (b) (5,12)
   (c) (6,6)
3. In the following assume $\theta \leq \frac{\pi}{2}$.
   (a) If $\sin \theta = \frac{5}{13}$, what is $\cos \theta$?
   (b) If $\cos \theta = \frac{1}{2}$, what is $\sin \theta$?
   (c) If $\sin \theta = \cos \theta$, what is $\sin \theta$? What is θ?
4. Show that for any right triangle with sides a, b and c as in Fig. 10.9,

$$\sin \theta = \frac{a}{c} \qquad \cos \theta = \frac{b}{c}$$

Fig. 10.9



5. Fill in Table 10.1 indicating the sine and cosine for each of the angles θ given in the table.

TABLE 10.1

θ	0	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$	2π
$\sin \theta$					
$\cos \theta$					

6. From the graph in Fig. 10.8(b) determine the following values, and then check your results with the sine table in the Appendix.
   - (a) $\sin 40^\circ$
   - (b) $\sin \frac{\pi}{12}$
   - (c) $\sin 70^\circ$

7. Using a unit circle, draw the graph of $y = \cos\theta$ for $0 \leq \theta \leq 2\pi$.

8. A right triangle having an acute angle of $\frac{\pi}{4}$ is isoceles, and two right triangles having angles of $\frac{\pi}{6}$ may be put back to back to form an equilateral triangle with an altitude bisecting one of the angles. Use this information to construct, in standard position, each on a separate coordinate system, the following angles given in radians and determine their sine and cosine.
   - (a) $\frac{\pi}{3}$
   - (b) $\frac{\pi}{4}$
   - (c) $\frac{\pi}{6}$

9. A boat sails on a course N40°E for 10 miles from point A to point B. How far east and how far north is B from A?

10. A helicopter climbs at a steady angle until it is 200 m above a point on the ground that is 300 m from the point of takeoff. What is the angle of climb?

11. On a set of coordinate axes, construct any angle θ. From the definitions of the sine and cosine functions given in this section, show that for the θ you have chosen
    - (a) $\sin(\pi - \theta) = \sin\theta$
    - (b) $\sin(\pi + \theta) = -\sin\theta$

12. Repeat the directions for Question 11 and show that
    - (a) $\cos(\pi - \theta) = -\cos\theta$
    - (b) $\cos(\pi + \theta) = -\cos\theta$

13. List all the values of θ in the interval $0 \leq \theta \leq 360^\circ$ that have the same
    - (a) sine as 30°
    - (b) cosine as 30°

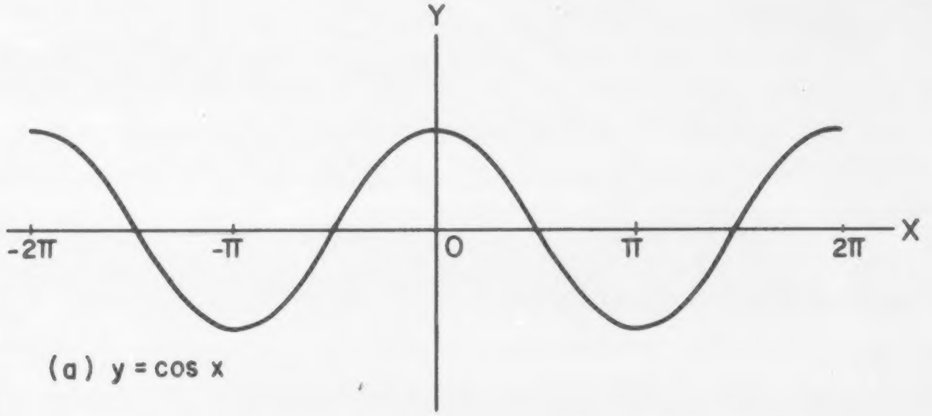
10.3 Symmetries and Antisymmetries

The graphs of $y = \cos x$ and $y = \sin x$ are shown in Fig. 10.10, where the negative angles are plotted to the left of the y axis. If we imagine folding the left side (negative portion) of the cosine graph in Fig. 10.10(a) over on top of the right side so that the fold is along the y axis, we observe that every point on one portion falls on the same point on the other. Thus, for any value of $\underline{x}$

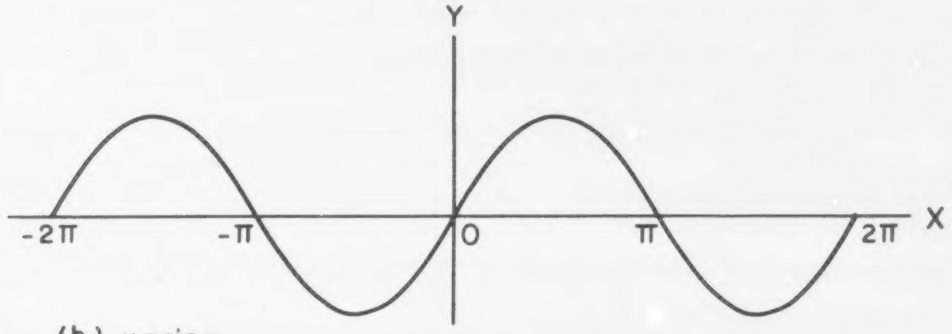
$$\cos(-x) = \cos x \tag{1}$$

In general, a function which has the property that $f(-x) = f(x)$ for all $\underline{x}$ in the domain of $f(x)$ is called an even function since the even power functions such as $f(x) = x^2$ or $f(x) = x^{-6}$ have this property. Hence the cosine function is an even function. Notice that the graph of an even function must be symmetric about the y axis.

Fig. 10.10



(a) $y = \cos x$



(b) $y = \sin x$

When the graph of $y = \sin x$ in Fig. 10.10(b) is folded over, the portions do not coincide. We observe, however, that

$$\sin(-x) = -\sin x \qquad (2)$$

Functions which have the property that $f(-x) = -f(x)$ are called odd functions since the odd power functions such as $f(x) = x$ or $f(x) = x^3$ have this property. Thus, the sine function is an odd function and its graph is said to be antisymmetric about the y axis.

Questions

1. Construct any negative obtuse angle θ. On the same set of axes construct the positive angle of the same magnitude. From the definitions of the sine and cosine, show that for the θ you have chosen

$$\sin(-\theta) = -\sin\theta$$

and

$$\cos(-\theta) = \cos\theta$$

2. Is the graph of $y = \sin x$ symmetric or antisymmetric about the vertical line through $x = \frac{\pi}{2}$? Explain.
3. Referring to Fig. 10.10(a), choose any point on the graph of $y = \cos x$ that lies to the left of the y axis. Locate the corresponding point to the right of the y axis. Are these two points symmetric or antisymmetric about the origin? Explain.

10.4 Periodicity

If an angle θ is in standard position and a second angle $(\theta + 2\pi)$ is in standard position, then the two angles have coincident terminal sides. Hence, they will have the same values for their sine and the same values for their cosine. We say, therefore, that the sine and cosine functions are periodic functions. We also say that both have the same period, 2π, because 2π is the smallest value added to an angle that makes the two angles co-terminal. We can express the periodicity of the sine and cosine functions as

$$\sin(\theta + 2\pi) = \sin\theta$$

and

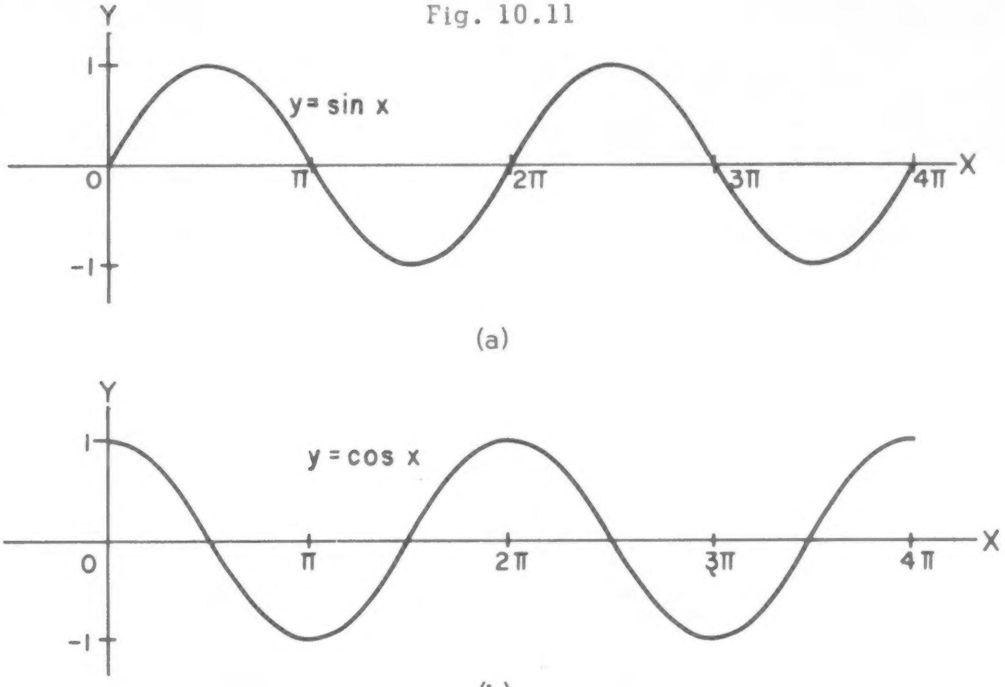
$$\cos(\theta + 2\pi) = \cos\theta$$

In general, a function is said to be periodic if there is a number $p \neq 0$ such that $f(x + p) = f(x)$ for all $\underline{x}$. The period of $f(x)$ is the smallest positive value of $\underline{p}$.

In Fig. 10.11 we have drawn the sine and cosine functions from 0 to 4π. In each case the graph begins to repeat itself at $\theta = 2\pi$, and had we continued plotting for larger values of θ, we would observe that after each interval of 2π radians the graphs would repeat.

Although $\sin\theta$ and $\cos\theta$ are periodic with respect to an angle, the world is full of other kinds of periodic functions. For example, the back-and-forth motion of an automobile piston and a pendulum in a clock, al-

Fig. 10.11



(a)

(b)

though not exactly sine functions, are periodic in time and this can be expressed as

$$f(t + T) = f(t)$$

where t is the time to reach a certain position y along the stroke of the piston and T is the period, the time to complete one back-and-forth motion. (See Fig. 10.12.)

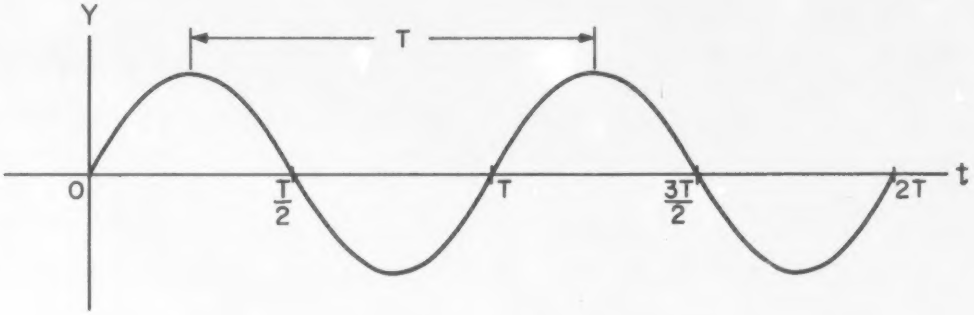


Fig. 10.12

Dirt roads often develop a repetitive pattern of small ridges and valleys running across the road. Water waves, particularly under controlled conditions as in a ripple tank, have repetitive patterns. These are but two of many examples of a periodic function of a length coordinate, which can be expressed as

$$f(x + L) = f(x)$$

where x, for example, is the distance as measured from some arbitrary point on the road and L is the distance between bumps (or valleys).

Of course, the periodic functions which occur in nature oscillate between many numerical values, not just between ± 1 as is the case with $y = \sin x$. Similarly, they need not have a period of 2π, and the values at $x = 0$ need not be either 0 or 1 (as with $y = \sin x$ and $y = \cos x$).

The graphs of many periodic functions are far from the shape of a sine curve. The position of point A on the movable pin, shown in

Fig. 10.13(a), as a function of the angle through which the cam is rotated is an example of a periodic function whose graph is not a sine or cosine curve. The graph of this displacement as a function of the angle x is shown in Fig. 10.13(b).

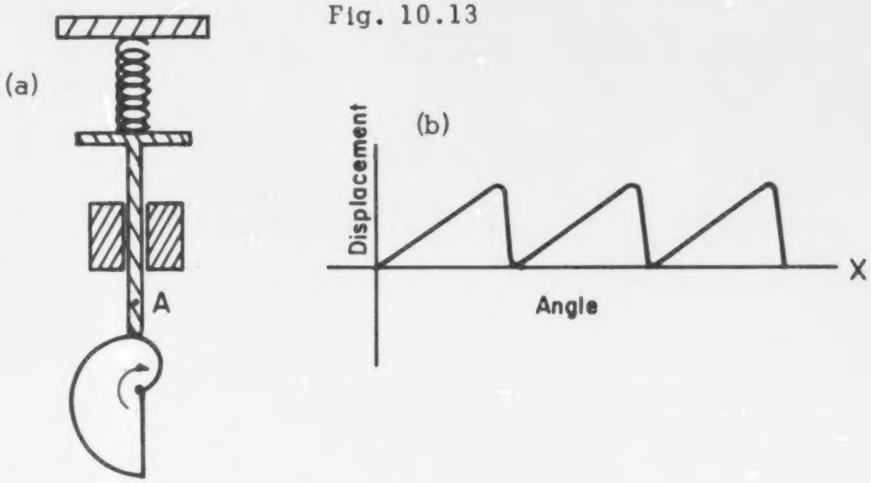


Fig. 10.13

Questions

1. Sketch the following angles on a coordinate system, and then express each in terms of a function of an acute angle.
   - (a) $\sin 120^\circ$
   - (b) $\sin 245^\circ$
   - (c) $\sin \frac{7\pi}{4}$
   - (d) $\cos 100^\circ$
   - (e) $\cos 320^\circ$
   - (f) $\cos \frac{4\pi}{3}$
2. Find the following angles by first expressing them as a function of an acute angle.
   - (a) $\sin(-610^\circ)$
   - (b) $\sin \frac{5\pi}{2}$
   - (c) $\sin 400^\circ$
   - (d) $\cos(-520^\circ)$
   - (e) $\cos 460^\circ$
3. What is the period of the function graphed in Fig. 10.13?

10.5 Homomorphic Trigonometric Functions

Once we have the graphs of the functions $y = \sin x$ and $y = \cos x$, we can examine curves homomorphic to them. For example, consider the function

$$y - y_0 = \sin x$$
$$y = \sin x + y_0$$

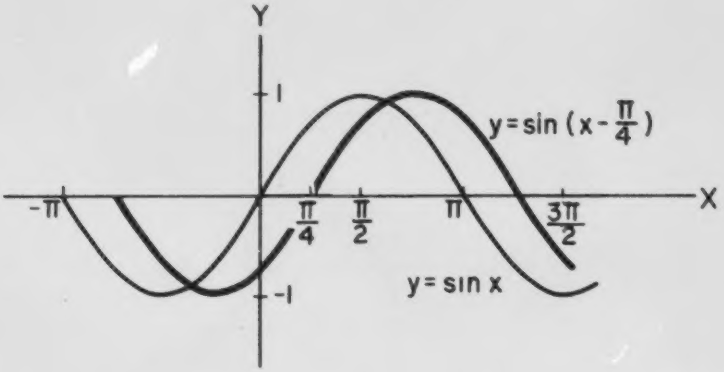
You will recall from Section 7.2 that the y_0 part of this function is an additive constant; it only moves the graph up or down, so the function oscillates between $y = y_0 \pm 1$.

Now consider the function

$$y = \sin (x - \frac{\pi}{4})$$

When $x = 0$ in this function, $y = \sin (-\frac{\pi}{4}) = -0.71$, so the graph does not start at either 0 or 1. The graph of this function is illustrated by the heavy line in Fig. 10.14. As you can see, each point on the $y = \sin x$ curve is displaced to the right by an amount $\frac{\pi}{4}$. For the function $y = \sin (x + \frac{\pi}{2})$, each point on the $y = \sin x$ curve is shifted by an amount $\frac{\pi}{2}$ to the left, as illustrated by the heavy line in Fig. 10.15. Notice that in this case the graph of $\sin (x + \frac{\pi}{2})$ is the same as the graph $y = \cos x$ (see Fig. 10.10(a)). Hence,

Fig. 10.14

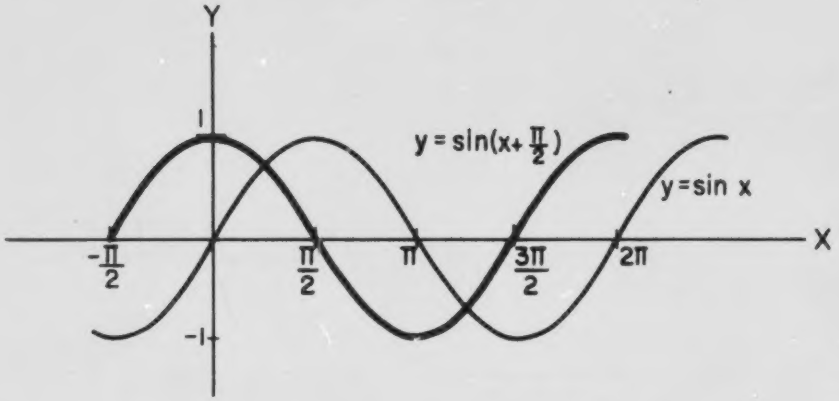


$$\cos x = \sin (x + \frac{\pi}{2})$$

Had we shifted the graph of $y = \cos x$ to the right by an amount $\frac{\pi}{2}$, and thus generated the graph $y = \cos(x - \frac{\pi}{2})$, then it would coincide with the graph of $y = \sin x$. That is,

$$\sin x = \cos(x - \frac{\pi}{2})$$

Fig. 10.15



It is because of this property of curves homomorphic to the sine and cosine functions that they are called co-functions and one is named "sine" and the other "cosine."

In general, then, we see that the graph of $y = \sin(x - x_o)$, where $(x_o > 0)$ is the same as the graph of $y = \sin x$ moved to the right by x_o units, and the graph of $y = \sin(x - x_o)$, where $(x_o < 0)$ is the graph of $y = \sin x$ moved to the left by x_o units. This shift of the graph to the left or to the right is often referred to as a change in phase and the number x_o is often called the phase angle.

Notice that if $x_o = \pm 2\pi$, the curve is shifted by exactly one period and coincides with $y = \sin x$. In other words, as stated in Section 10.4, $\sin(x + 2\pi) = \sin x$ or, more generally,

$$\sin(x + 2n\pi) = \sin x$$

for any integral value of n. Homomorphic curves given by different values of n exactly overlap.

Questions

1. Sketch the graphs of the following functions over the interval $-\pi \leq x \leq 3\pi$:
   (a) $y = \sin x$
   (b) $y = \sin (x + \frac{\pi}{3})$
   (c) $y = \sin (x - 60^{\circ})$

2. Sketch the graphs of the following functions over the interval $-\pi \leq x \leq 3\pi$
   (a) $y = \cos x$
   (b) $y = \cos (x + \frac{\pi}{4})$
   (c) $y = \cos (x - 30^{\circ})$

3. From the relation

$$\sin \theta = \cos (\theta - \frac{\pi}{2})$$

developed in this section, which is true for any value of θ, show that in particular, when $\theta + \phi = \frac{\pi}{2}$, then

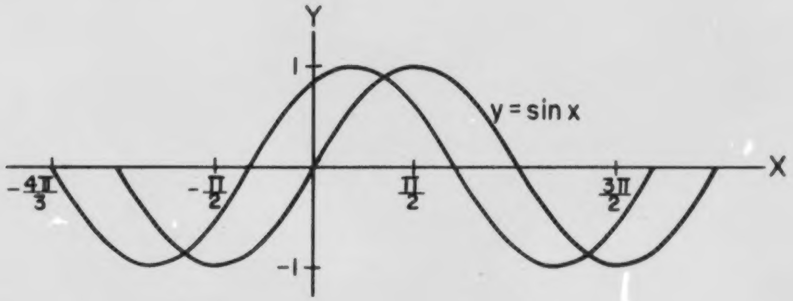
$$\sin \theta = \cos \phi$$

that is, co-functions of complementary angles are equal.

4. Sketch the graph of $y = \sin (x + \frac{\pi}{4})$. On the same axes, sketch the graph that is homomorphic to it and has a phase angle of $\frac{\pi}{2}$. Is there more than one way to do this?

5. Write the equation of the graph homomorphic to $y = \sin x$ shown in Fig. 10.16.

Fig. 10.16



6. Show that if two functions are homomorphic, they have the same period, T. That is, prove that if

$$\sin (x + T) = \sin x$$

then

$$\sin ([x - x_0] + T) = \sin (x - x_0)$$

10.6 The Functions y = A sin (kx) and y = cos (kx)

Just as with other functions, we can multiply a trigonometric function by a constant. Thus, the graph of $y = 3 \cos x$ is like that of $y = \cos x$ except that each y coordinate is three times greater (Fig. 10.17). The absolute value of the constant coefficient A in $y = A \sin x$ and $y = A \cos x$ is called the amplitude of the function.*

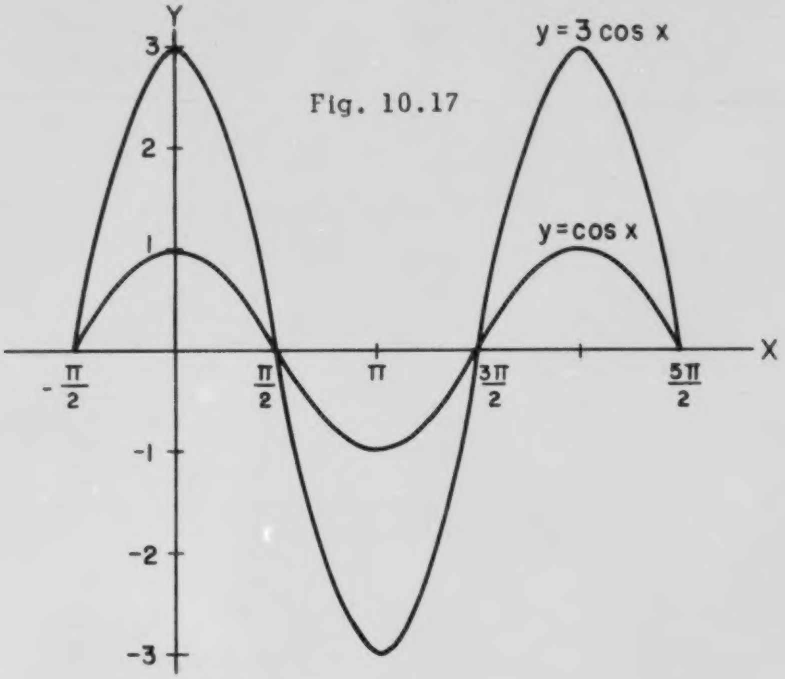


Fig. 10.17

We next examine the graphs of functions of the form $y = \sin (kx)$ for various values of k. We begin with $k = 2$, that is, $y = \sin 2x$, and make a table of values. From this table (Table 10.2) we have drawn a graph of the function $y = \sin 2x$. As you can see from Fig. 10.18, the function has a period $T = \pi$, half the value of the period of the function $y = \sin x$.

---

*More generally, the amplitude of a function is defined as one-half the difference of the maximum and minimum displacements from the zero position.

Next, examine the graph of the function $y = \sin\frac{1}{2}x$ shown in Fig. 10.19. Here, $y = \sin\frac{1}{2}x$ goes through a complete period as x goes from 0 to 4π. So the period of $y = \sin\frac{1}{2}x$ is $T = 4\pi$.

In general, the value of k in the functions $y = \sin(kx)$ and $y = \cos(kx)$ determines the period T:

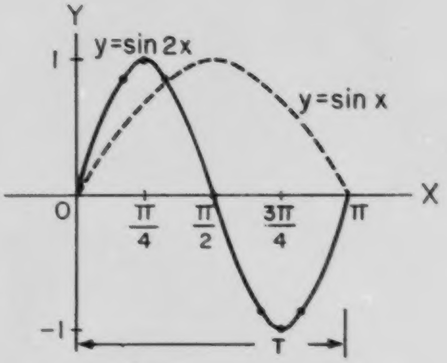
$$T = \frac{2\pi}{k}$$

where k is the number of periods in an interval of length 2π.

TABLE 10.2

x	$2x$	$\sin 2x$
0	0	0
$\frac{\pi}{6}$	$\frac{\pi}{3}$	$\frac{\sqrt{3}}{2}$
$\frac{\pi}{4}$	$\frac{\pi}{2}$	1
$\frac{\pi}{3}$	$\frac{2\pi}{3}$	$\frac{\sqrt{3}}{2}$
$\frac{\pi}{2}$	π	0
$\frac{2\pi}{3}$	$\frac{4\pi}{3}$	$-\frac{\sqrt{3}}{2}$
$\frac{3\pi}{4}$	$\frac{3\pi}{2}$	-1
$\frac{5\pi}{6}$	$\frac{5\pi}{3}$	$-\frac{\sqrt{3}}{2}$
π	2π	0

Fig. 10.18



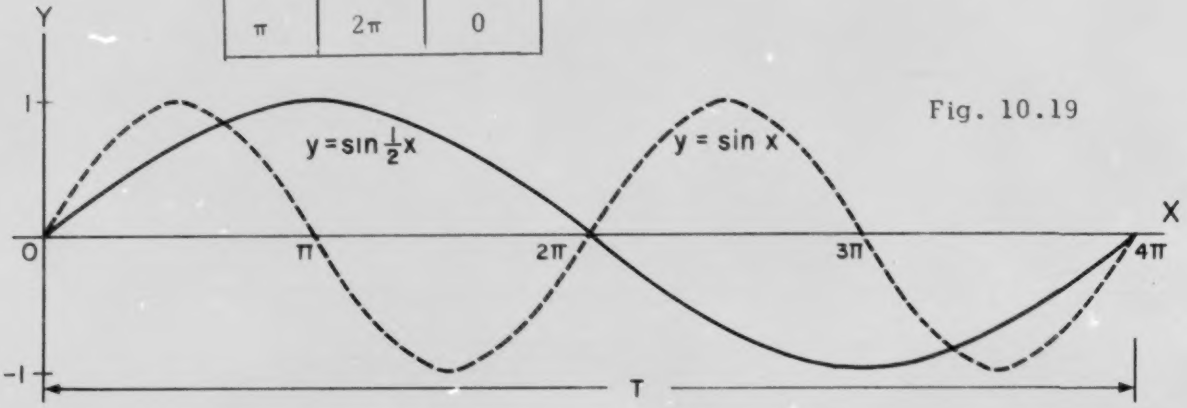


Fig. 10.19

Questions

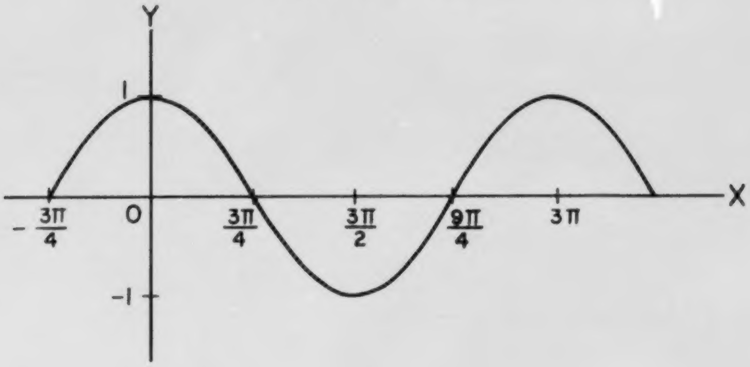
1. Sketch the following graphs on the same set of axes and discuss where the graphs are increasing, where they are decreasing, and where they reach their maxima and minima. Determine the period of each.
   (a) $y = \cos 2x$
   (b) $y = \cos 3x$
   (c) $y = \cos \frac{1}{2}x$

2. Sketch the following graphs on the same set of axes and discuss where the graphs are increasing, decreasing, and where they reach their maxima and minima.
   (a) $y = 2 \cos x$
   (b) $y = 3 \cos x$
   (c) $y = \frac{1}{2}\cos x$

3. Repeat the directions for Question 2 for the following functions.
   (a) $y = \cos (x + \frac{\pi}{2})$
   (b) $y = \cos (x - \frac{\pi}{2})$
   (c) $y = \cos (x + \pi)$

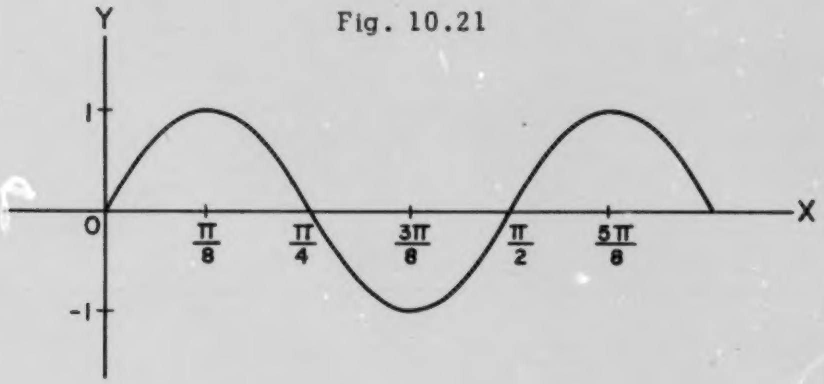
4. The graph shown in Fig. 10.20 is a cosine function. What is its equation?

Fig. 10.20



5. The graph shown in Fig. 10.21 is a sine function. What is its equation?

Fig. 10.21



6. Discuss how the value of k will affect the graph of
   - (a) $y = k \cos x$
   - (b) $y = \cos (kx)$
   - (c) $y = \cos (x + k)$

10.7 The Functions $A \sin k (x - x_0)$ and $y = A \cos k (x - x_0)$

Putting together the ideas of the last two sections, we can write general forms for both the sine and cosine functions. They are

$$y = A \sin k (x - x_0) + y_0$$
$$y = A \cos k (x - x_0) + y_0$$

All we need to do to sketch either of these two functions is to sketch the corresponding equation $y = \sin x$ or $y = \cos x$ and adjust the y axis, the x axis, the amplitude, and the period, as necessary.

We will illustrate this procedure with an example. To sketch the function

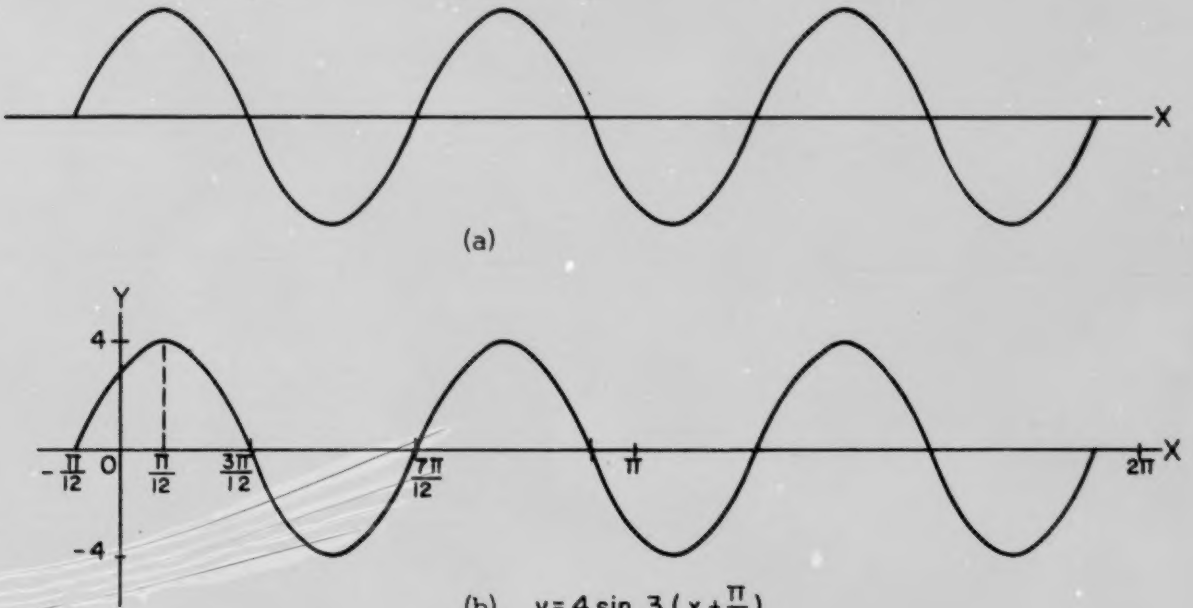
$$y = 4 \sin 3 (x + \frac{\pi}{12})$$

we first note that the amplitude is 4. This says to stretch the vertical axis by a factor of 4, that is, the maximum is +4 and the minimum is -4.

The coefficient 3 tells us there are 3 periods in any 2π interval. So we sketch 3 periods of sin 3x, not yet graduating the x axis (Fig. 10.22(a)).

To adjust the y axis we note that the sine is zero when the angle $3(x + \frac{\pi}{12}) = 0$, that is, at $x = -\frac{\pi}{12}$. Thus the beginning of the sine period is at $x = -\frac{\pi}{12}$. We can now draw the y axis and appropriately graduate the x axis, remembering that one period is $\frac{2\pi}{3}$ units long, as shown in Fig. 10.22(b).

Fig. 10.22



(a)

(b) $y = 4 \sin 3(x + \frac{\pi}{12})$

Questions

1. Sketch the graphs of
   - (a) $y = 4 \sin (x - \frac{\pi}{2})$
   - (b) $y = 2 \sin (x + \pi)$
   - (c) $y = 3 \cos (x - \frac{\pi}{4})$
2. Sketch the graphs of
   - (a) $y = 2 \sin (x + \frac{\pi}{4}) - 3$
   - (b) $y = 3 [\cos (x - \frac{\pi}{4}) + 1]$

3. Sketch on the same set of axes the graph of $y = -2 \sin x$ and $y = 2 \sin (x - \pi)$. What does the sign of the coefficient 2 tell you about the phase angle of the first function?

4. Sketch the following graphs and discuss each with regard to maxima, minima, zeros, period, and phase angle.
   - (a) $y = 2 \cos (x + \frac{\pi}{8})$
   - (b) $y = 4 \cos (2x - \pi)$
   - (c) $y = \sin 2 (x + \frac{\pi}{4})$
   - (d) $y = \sin (3 [x - \frac{\pi}{6}])$

10.8 Recognizing Trigonometric Functions from a Graph

We shall now use the procedure of the last section in reverse. That is, given a graph of a periodic function, how can we find out if it is expressible in terms of a sine or cosine function? For example, can the graph in Fig. 10.23, which is a periodic function, be expressed in the form

$$y = A \cos k (x - x_0)?$$

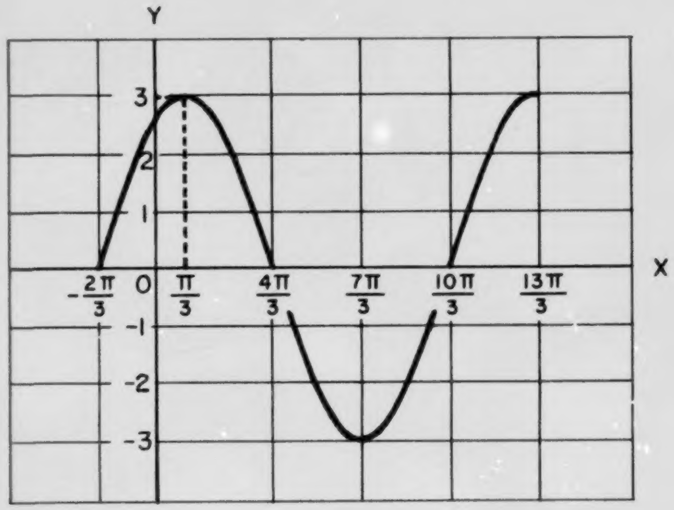


Fig. 10.23

Since the function is periodic, we can find k by noting that one period on the graph extends from $-\frac{2\pi}{3}$ to $\frac{10\pi}{3}$, or a total length of 4π. Therefore, in a 2π interval there is one-half period. So $k = \frac{1}{2}$. We also note from the graph that $A = 3$. Hence, the function describing the graph in Fig. 10.23 has the form

$$y = 3 \cos \frac{1}{2}(x - x_o)$$

where it remains to find the value of x_o. Notice that the first maximum occurs at $\frac{\pi}{3}$, so this must be the phase angle. The complete equation is now written as

$$y = 3 \cos \frac{1}{2}(x - \frac{\pi}{3})$$

We are not finished, however. It is not sufficient to conclude that this equation is indeed the correct function just because the amplitude, period and phase angle are in agreement with those of the graph. It remains to test intermediate values for x and y in the function to see if they are related by a cosine function.

The test we use is like the one we have used before to test for parabolas, hyperbolas, etc. In the case of a curve which you suspected was a parabola of the form $y - n = a(x - m)^2$ you plotted $y - n$ as a function of the quantity $(x - m)^2$. If the graph you obtained was a straight line you concluded that the function $y - n = a(x - m)^2$ was indeed the function describing the curve in question.

In the case of the function we are now considering, if we graph y as a function of $\cos \frac{1}{2}(x - \frac{\pi}{3})$ and get a straight line we know we have a cosine function. Table 10.3 shows the values of x and y as read from the graph in Fig. 10.23 and the corresponding values of $\cos \frac{1}{2}(x - \frac{\pi}{3})$ up to $x = \frac{\pi}{3}$. Figure 10.24 was made from this table. The size of the error bars in Fig. 10.24 is based on an estimate of the errors in reading the values of y from the graph in Fig. 10.23 and in the plotting of Fig. 10.24. As you can see, a straight line, whose slope is 3, passes through all the error bars.

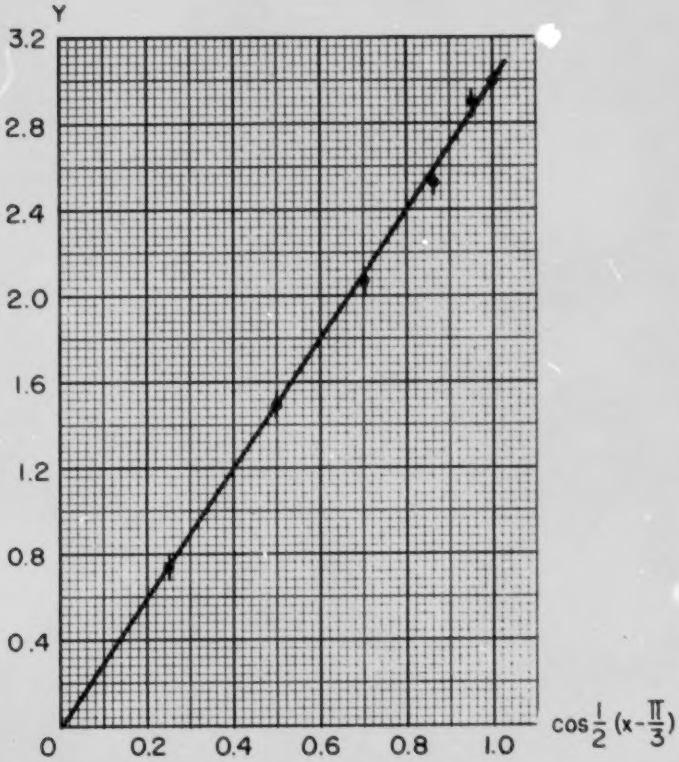
Since the curve in Fig. 10.23 is very smooth between the plotted values, we feel confident that all other intermediate points we might care to plot on the graph in Fig. 10.24 would fall within the error bars.

We now test the rest of a complete period by checking for symmetry about lines perpendicular to the x axis that pass through the maxima and minima and for antisymmetry about a perpendicular line through $x = \frac{4\pi}{3}$. From this and the graph in Fig. 10.24 we conclude that the curve in Fig. 10.23 is indeed described up to $x = \frac{\pi}{3}$ by the function $y = 3 \cos \frac{1}{2}(x - \frac{\pi}{3})$ to within the errors of reading and plotting.

TABLE 10.3

x	$\cos \frac{1}{2}(x - \frac{\pi}{3})$	y
$-\frac{2\pi}{3}$	0	0
$-\frac{\pi}{2}$	0.259	0.78
$-\frac{\pi}{3}$	0.500	1.48
$-\frac{\pi}{6}$	0.707	2.08
0	0.866	2.58
$+\frac{\pi}{6}$	0.966	2.90
$+\frac{\pi}{3}$	1.00	3.00

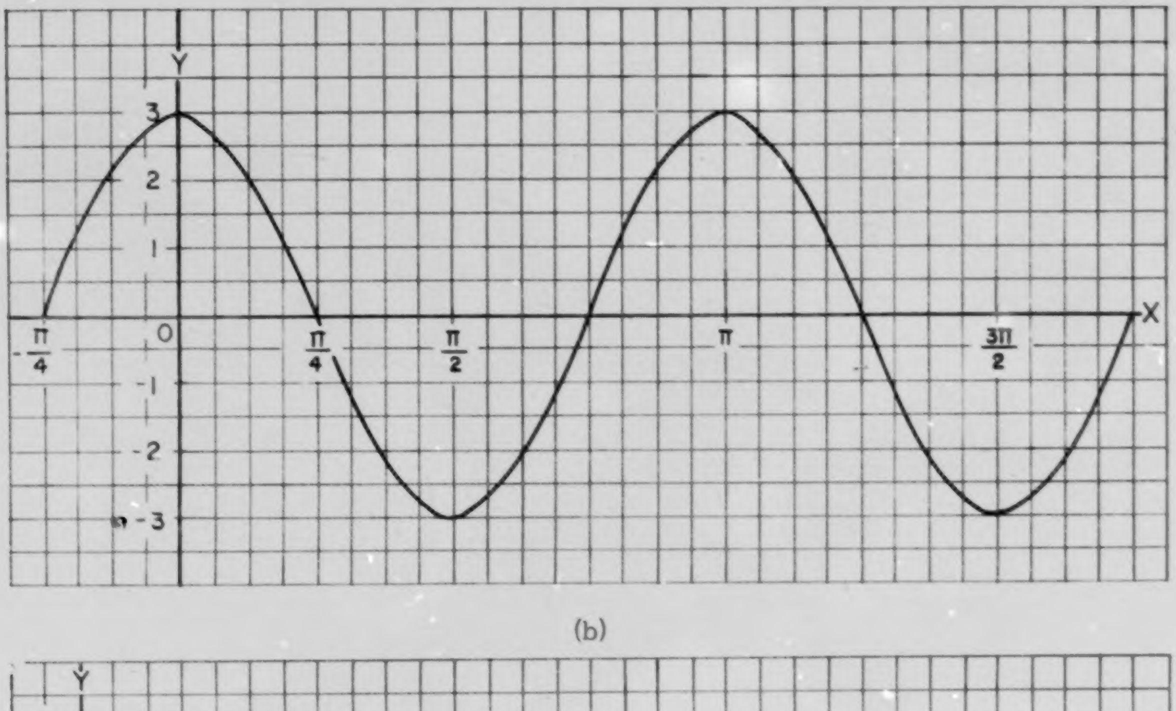
Fig. 10.24



Questions

1. Determine the amplitude, period, and phase angle of the graphs in Fig. 10.25 and then write their equations. (Is there more than

Fig. 10.25 (a)

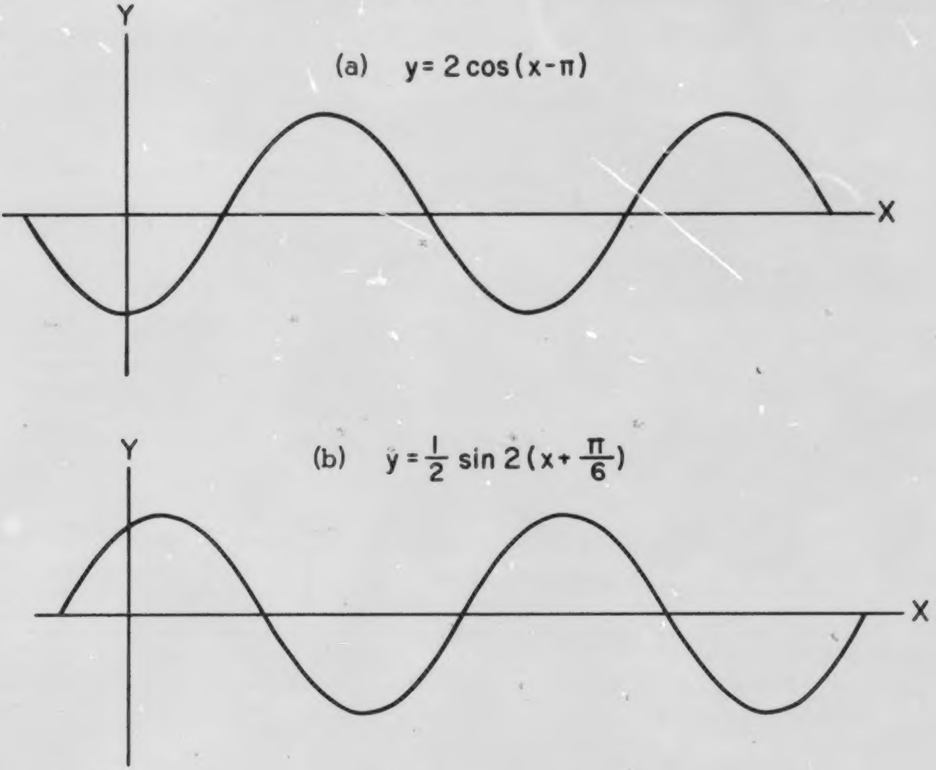


one answer to this question?) Choose values of x and check to see if they satisfy the equations you have written.

2. Could we have expressed the equation of the graph in Fig. 10.23 as a sine function? What would this equation be? Check your result.

3. Two functions and their graphs are shown in Fig. 10.26. Label the x and y axes at the zeros of the functions.

Fig. 10.26

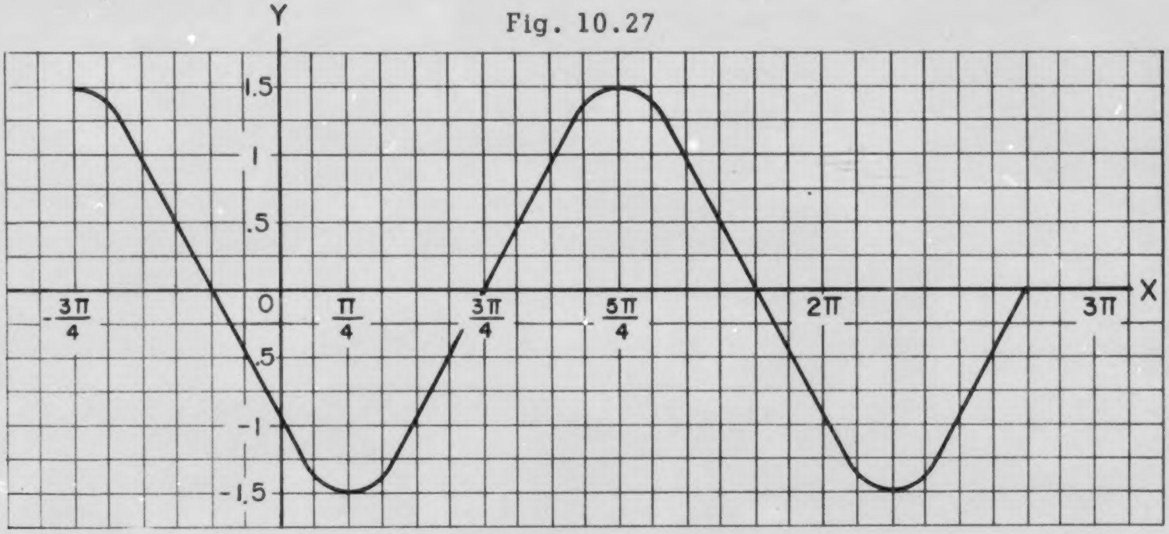


4. How would your answer to Question 4, Section 10.6 change if the graph given was a sine function?

5. How would your answer to Question 5, Section 10.6 change if the given graph was a cosine function?

6. Determine if the graph in Fig. 10.27 represents sine or cosine functions by writing the sine and cosine functions that fit the maximum and minimum points on the graph and then checking some intermediate points.

Fig. 10.27



7. Plot the graph of the data in the following table and write the equation of the sine function it describes. (Graduate the x axis in integral numbers of radians.)

x (rad)	y
0	-0.90
0.30	-1.42
0.46	-1.50
0.60	-1.45
0.90	-0.97
1.10	-0.45
1.20	-0.15
1.40	0.45
1.50	0.72
1.57	0.90
1.80	1.34
2.03	1.50
2.20	1.42

8. The relationship between the size of an angle and the ratio of certain sides of the right triangle that contains the angle can be expressed by functions other than the sine and cosine. With the angle in standard position (Fig. 10.28), we now define the tangent function and the cotangent function such that

$$\tan\theta = \frac{y}{x}$$

and

$$\cot\theta = \frac{x}{y}$$

(a) Are the tangent and cotangent really cofunctions?

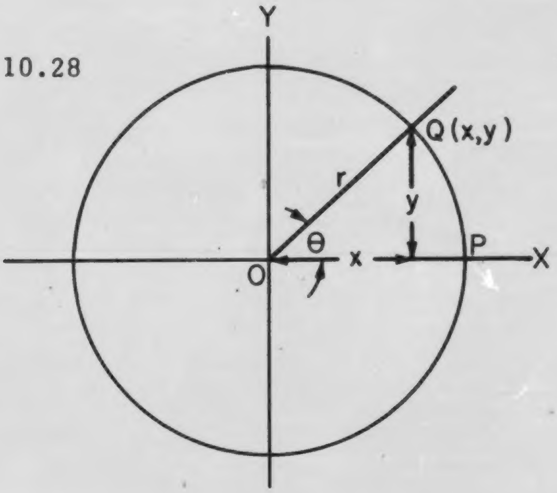
(b) For what values of θ is the tangent function defined?

(c) For what values of θ is the cotangent function defined?

(d) Draw the graphs of the tangent and cotangent functions from $x = 0$, to $x = 2\pi$.

(e) Are the tangent and cotangent functions periodic?

Fig. 10.28



9. (a) Is the tangent function odd or even?

(b) Is the cotangent function odd or even?

10. Prove the following relations by constructing the angle θ on coordinate axes and applying the definition of the tangent.

(a) $\tan(-\theta) = -\tan\theta$

(b) $\tan(\pi - \theta) = -\tan\theta$

(c) $\tan(\pi + \theta) = \tan\theta$

11. What values of θ in the interval $0 \leq \theta \leq 2\pi$ have the same tangent as
    (a) 45°
    (b) -45°

12. Sketch the following angles on a coordinate system, and then express each in terms of a function of an acute angle.
    (a) $\tan 120^\circ$
    (b) $\tan 245^\circ$
    (c) $\tan \frac{7\pi}{4}$
    (d) $\cot 100^\circ$
    (e) $\cot 320^\circ$
    (f) $\cot \frac{4\pi}{3}$

13. The ramp leading up to a bridge makes an angle of 5° with the horizontal. How much vertical rise is there in a horizontal distance of 10 meters?

10.9 Qualitative Observations on the Derivatives of $\sin x$ and $\cos x$

The goal of the remainder of this chapter is to find the derivative of the sine and cosine functions. Before doing this, however, we can draw some conclusions about the nature of these derivatives from the properties of the sine and cosine functions themselves. In other words, we can make some "ball-park" predictions as to what type of function we expect to turn up as the derivative of $\sin x$ or $\cos x$.

For example, recall that the sine function is periodic with a period of 2π. This means that the graph of $\sin x$ for $\underline{x}$ between 0 and 2π coincides with the graph of $\sin x$ for $\underline{x}$ between 2π and 4π. An immediate conclusion from this is that the graph of the slope function for $\sin x$ from 0 to 2π must also coincide with the graph of the slope function for $\sin x$ from 2π to 4π. More generally, we can conclude that the derivative of the sine function must be a periodic function with period 2π.

By inspecting the graph of $y = \sin x$ in Fig. 10.10(b), we observe that the largest value of the slope function occurs at the origin (and at multiples of 2π) and seems to be equal to 1. In fact, it appears that the range of the derivative of the sine function is roughly the interval -1 to 1.

A final observation is that while $\sin x$ is an odd function its derivative is not. For example, the slope of the graph of $\sin x$ at $x = \pi$ is obviously negative as is the slope at $x = -\pi$. Hence, the derivative of the sine function cannot satisfy the equation $f(-x) = -f(x)$ which characterizes odd functions. Note, however, that the slope of the graph of $\sin x$ at $x = -\frac{\pi}{2}$ appears to be the same as the slope at $x = \frac{\pi}{2}$; namely, zero. It is quite likely from the graph of $\sin x$ that the derivative of the sine function is an even function.

In summary, we expect the derivative of the sine function to (a) be periodic with period 2π, (b) have a range from -1 to 1, and (c) be an even function.

To find an exact analytic expression for the derivative of $\sin x$ it is natural to start with the delta process. For the derivative of $\sin x$ we write

$$[\sin x]' = \lim_{\Delta x \to 0} \frac{1}{\Delta x} (\sin (x + \Delta x) - \sin x) \tag{3}$$

There appears to be no way in which we can cancel out the Δx's. However, it is possible to derive the addition formula for the sine functions (a formula for the sine of the sum of two angles) to replace the first term in Equation (3). We can then write Equation (3) in a form from which we can find the limit as Δx approaches zero. Then, by a similar process, we can find $[\cos x]'$.

Questions

1. What are the largest and smallest values that you expect for the derivative of $\cos x$?
2. $\cos x$ is an even function. Is its derivative an even function? An odd function?

3. Verify from the graph of the cosine function that its derivative is periodic with period 2π.
4. If $f(x)$ is a periodic function, not necessarily the sine or cosine function, must $f'(x)$ also be a periodic function? Why or why not?
5. If $f(x)$ is a periodic function, must its antiderivative be periodic? Why or why not?

10.10 The Addition Formula for the Sine Function

We shall derive the addition formula for the sine function, where the two angles α and β are both acute (less than 90°) as shown in Fig. 10.29.

Fig. 10.29

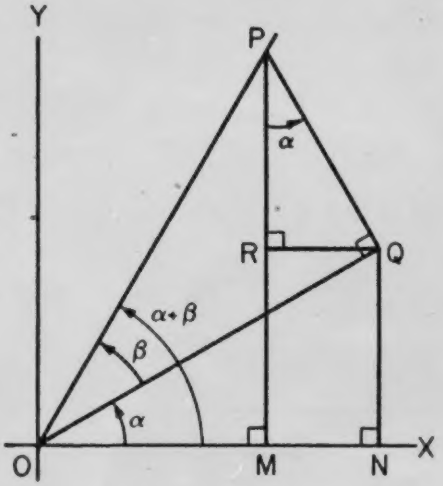


Figure 10.29 is constructed as follows: The angle α is drawn in standard position with its terminal side along OQ, and β is drawn with OQ as its initial side and OP as its terminal side. The figure is completed by dropping perpendiculars PM, QN, PQ, and QR. The two angles labeled α are equal because their sides are mutually perpendicular. In the figure

$$MP = NQ + PR$$

$$MP = OQ \sin\alpha + QP \cos\alpha$$

Dividing the second expression by OP,

$$\frac{MP}{OP} = \frac{OQ}{OP}\sin\alpha + \frac{QP}{OP}\cos\alpha$$

and, therefore,

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta \qquad (4)$$

This is the addition formula for the sine function. Although our construction holds only for acute angles, Equation (4) is true for any values of α and β, including negative angles.

Questions

1. Use Fig. 10.30 to show that $\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$ if $\alpha+\beta$ is obtuse, starting with the relation OM = MN - ON. (Hint: use the relation $\sin(\pi-\theta) = \sin\theta$.)

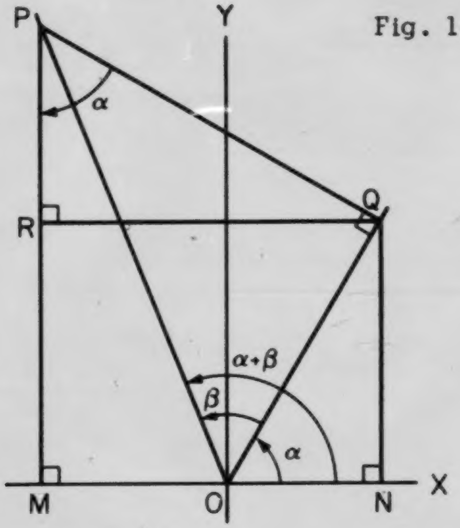


Fig. 10.30

2. Show that $\sin(\alpha-\beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta$ by substituting $\theta = -\beta$ in Equation (2).
3. Is the addition formula for $\sin(\alpha+\beta)$ true when one of the angles is equal to zero?
4. By letting $\alpha = \beta$ in the addition formula, derive the expression for $\sin(2\alpha)$.

10.11 The Derivative of sin x

We are now ready to apply the addition formula for the sine function to sin (x + Δx) in Equation (3):

$$\sin(x + \Delta x) = \sin x \cos \Delta x + \cos x \sin \Delta x$$

This changes the expression for [sin x]' to

$$[\sin x]' = \lim_{\Delta x \to 0} \frac{\sin x \cos \Delta x + \cos x \sin \Delta x - \sin x}{\Delta x}$$

and after rearranging:

$$[\sin x]' = \lim_{\Delta x \to 0} \left(\sin x \left(\frac{\cos \Delta x - 1}{\Delta x} \right) + \cos x \, \frac{\sin \Delta x}{\Delta x} \right) \tag{5}$$

Let us look at each of the two limits in Equation (5) separately. The first limit, $\lim\limits_{\Delta x \to 0} \frac{\cos \Delta x - 1}{\Delta x}$ can be looked at as

$$\lim_{\Delta x \to 0} \frac{\cos(0 + \Delta x) - \cos 0}{\Delta x}$$

This is the derivative of cos x at x = 0. A simple interpretation of the graph of y = cos x (Fig. 10.31) shows that the tangent to the curve at x = 0 has a slope equal to zero, therefore

$$\lim_{\Delta x \to 0} \frac{\cos \Delta x - 1}{\Delta x} = 0 \tag{6}$$

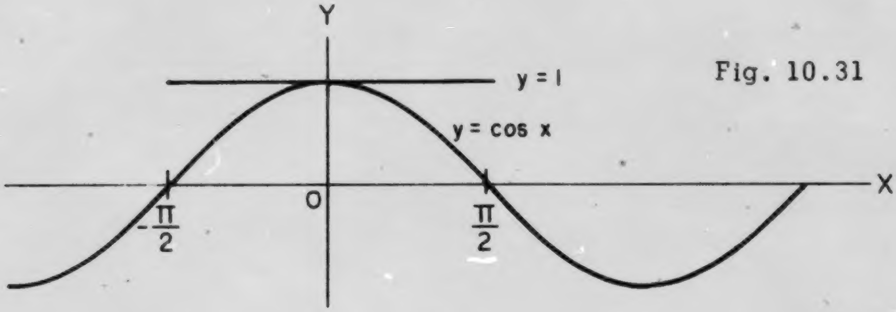


Fig. 10.31

The second limit can also be rewritten in a similar way:

$$\lim_{\Delta x \to 0} \frac{\sin \Delta x}{\Delta x} = \lim_{\Delta x \to 0} \frac{\sin(0 + \Delta x) - \sin 0}{\Delta x}$$

The right hand side of the last equation is the derivative of $\sin x$ at $x = 0$. In Fig. 10.32 the graph of $y = \sin x$ shows the slope of the tangent to the curve at $x = 0$ is 1. Thus

$$\lim_{\Delta x \to 0} \frac{\sin \Delta x}{\Delta x} = 1 \tag{7}$$

Substituting Equations (6) and (7) into Equation (5) yields

$$[\sin x]' = \cos x \tag{8}$$

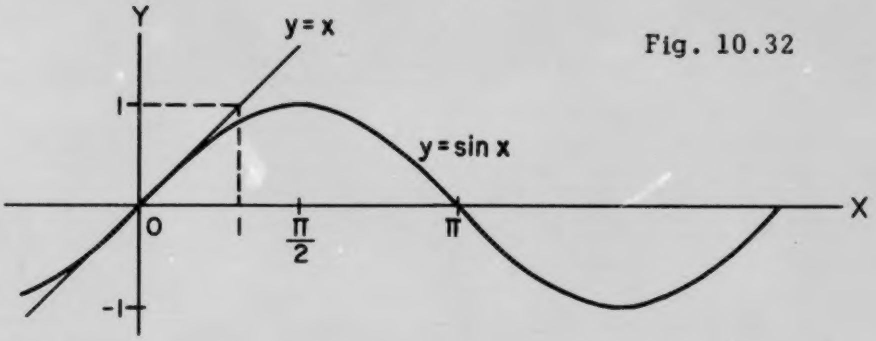


Fig. 10.32

Questions

1. (a) What is the period of the function $\sin 2x$?

   (b) What do you expect the period of $[\sin 2x]'$ to be?

2. Find the derivative of $\sin kx$, where k is a constant, using the delta process. (The following equations will be useful:)

$$\lim_{\Delta x \to 0} \frac{\sin k\Delta x}{\Delta x} = \lim_{\Delta x \to 0} k \frac{\sin k\Delta x}{k\Delta x} = k \lim_{\Delta x \to 0} \frac{\sin k\Delta x}{k\Delta x}$$

3. (a) Given that $\sin\frac{\pi}{6} = 0.05$, how can you use your knowledge of the derivative of $\sin x$ to find an approximate value for $\sin\frac{\pi}{5}$? (Hint: See Section 8.8.)

   (b) From your knowledge of $\sin 45^\circ$, find an approximate value for $\sin 48^\circ$ and $\sin 42^\circ$.

   (c) Compare your results with the values given in a table.

4. Give an approximation for $\sin x$ near $x = 0$

   (a) when x is expressed in radians.

   (b) when x is expressed in degrees.

5. (a) What is the family of antiderivatives of $\cos x$?

(b) What is the antiderivative of $f(x) = 3\cos x$ that satisfies the initial condition $F(0) = 5$?

6. Calculate the following integrals

(a) $\displaystyle\int_0^{\frac{\pi}{2}} \cos x\, dx$

(b) $\displaystyle\int_0^{\pi} \cos x\, dx$

(c) $\displaystyle\int_0^{2\pi} \cos x\, dx$

(d) What is the geometric interpretation of these integrals?

10.12 The Derivative of cos x

Finding the derivative of $\cos x$ involves steps similar to those used in the preceding two sections for finding the derivative of $\sin x$. To be able to apply the delta process requires that we know how to express $\cos(x + \Delta x)$ in terms of the sine and cosine of x and Δx.

Figure 10.29 (reproduced here as Fig. 10.33) will serve to find the general expression for the cosine of the sum of two angles α and β. From Fig. 10.31

$$OM = ON - MN$$

$$OM = OQ\cos\alpha - PQ\sin\alpha$$

Dividing both sides by OP:

$$\frac{OM}{OP} = \frac{OQ}{OP}\cos\alpha - \frac{PQ}{OP}\sin\alpha$$

Thus,

$$\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta \qquad (9)$$

holds when both α and β are acute. As in the case of the formula for $\sin(\alpha + \beta)$ Equation (9) holds for all positive and negative values of α and β.

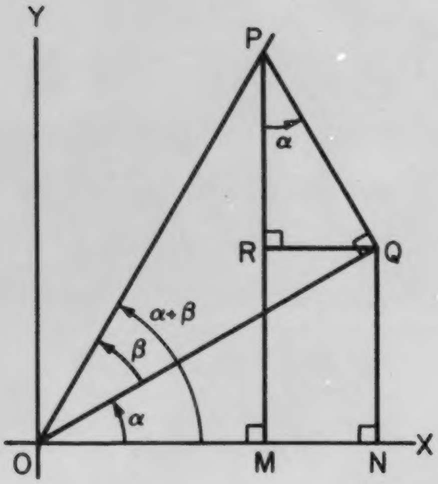


Fig. 10.33

Now that we have an expression for cos (α + β) we can use the delta process to find [cos x]'. The derivation is much like that of [sin x]'.

$$[\cos x]' = \lim_{\Delta x \to 0} \frac{\cos(x + \Delta x) - \cos x}{\Delta x}$$

From the addition formula,

$$[\cos x]' = \lim_{\Delta x \to 0} \frac{\cos x \cos \Delta x - \sin x \sin \Delta x - \cos x}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \cos x \left(\frac{\cos \Delta x - 1}{\Delta x}\right) - \lim_{\Delta x \to 0} \sin x \left(\frac{\sin \Delta x}{\Delta x}\right)$$

These are the same limits that appeared in Equations (6) and (7). Hence,

$$[\cos x]' = -\sin x \qquad (10)$$

Questions

1. Find the derivative of cos kx.
2. Give an approximation for cos x near x = 0 using the approach of Section 8.8.
3. What is the family of antiderivatives of sin x?
4. Find an antiderivative of sin kx.

5. Find the antiderivative of $f(x) = 10 \sin x$ that satisfies the initial condition $F(0) = 0$.

The functions $\sin x$ and $\cos x$ have the property that $[\sin x]' = \cos x$ and $[\cos x]' = -\sin x$. Consider the two functions

$$g_1(x) = \frac{1}{2}(e^x + e^{-x}) \text{ and } g_2(x) = \frac{1}{2}(e^x - e^{-x})$$

6. Is there a similar relationship between these functions and their derivatives?

7. Suppose a mass tied to the end of a spring oscillates up and down (Fig. 10.34). Its vertical position as a function of time is given by $x = 5 \cos 3\pi t$.

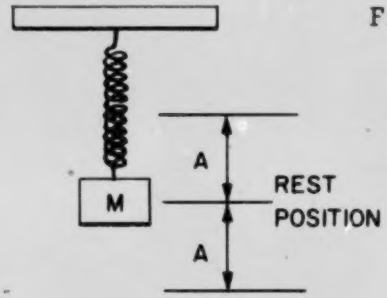


Fig. 10.34

(a) At what times is the mass at the (i) highest (ii) lowest point?

(b) The velocity of the mass is given by the derivative of the position with respect to time. Find the velocity as a function of time.

(c) What is the velocity at the highest and lowest point? Is this surprising?

APPLIED MATHEMATICS I

Appendix 1: Algebraic Manipulations

1. Some Properties of Numbers

Is it possible to simplify the algebraic expression $\dfrac{a^2 - \dfrac{a+3}{b}}{a}$ and, if so, where does one start? We must remember that the algebraic expressions we have been working with have involved numbers and variables that stand for numbers. Therefore, we can handle an algebraic expression as we would handle any expression involving numbers. Let us review some properties of numbers that will aid in the simplification of algebraic expressions.

Three important properties of the number system are:

1. The Associative Property
   - (a) for addition $a + (b + c) = (a + b) + c$
   - (b) for multiplication $a\,(bc) = (ab)\,c$
2. The Commutative Property
   - (a) for addition $a + b = b + a$
   - (b) for multiplication $ab = ba$
3. The Distributive Property
   - (a) $a\,(b + c) = ab + ac$, or $(a + b)\,c = ac + bc$

Also, we have the definition of subtraction,

4. $\boxed{a - b = a + (-b)}$

and, finally, some important results of elementary algebra.

$$\boxed{\begin{array}{l} 5.\ -a = -1 \cdot a \\ 6.\ (-a) \cdot b = -(ab) \\ 7.\ (-a) \cdot (-b) = a \cdot b \\ 8.\ -(a + b) = -a - b \end{array}}$$

The associative property allows us to remove or insert parentheses between the terms of an algebraic sum (or product). For example, using the associative property we can write $(3x + 2y) + (5x + 6)$ as $3x + 2y + 5x + 6$; and

$(3x)\ (4x)$ as $3 \cdot x \cdot 4 \cdot x$; and

$(5x + 3y) + 2(x - y)$ as $5x + 3y + 2(x - y)$; and

$(ab)\ (c + d)$ as $ab(c + d)$

Notice, however, that in an expression like $a - (b + c + d)$ we must be careful because we are not dealing strictly with a sum. If we rewrite the expression, using property 8, as $a - b - c - d$ and rewrite this as $a + (-b) + (-c) + (-d)$, using property 4, we may now group the terms as we please (e.g., $(-b) + (-d) + (-c) + a$).

The commutative property allows us to change the order of the terms of an algebraic sum (or product). For example, we can write:

$3x + 2y + 5x + 6$ as $3x + 5x + 2y + 6$; and

$3 \cdot x \cdot 4 \cdot x$ as $3 \cdot 4 \cdot x \cdot x$; and

$3\ (x + 3) + 6$ as $6 + 3\ (x + 3)$; and

$(a + b) \cdot (c + d)$ as $(c + d) \cdot (a + b)$

The distributive property is the one number property that ties multiplication and addition together. The distributive property permits us to write:

$3x + 5x$ as $(3 + 5)\ x$; and

$abc + ad$ as $a\ (bc + d)$; and

$(a - b)\ (a + b)$ as $(a - b)\ a + (a - b)\ b$; and

$u\ (s + t) + v\ (s + t)$ as $(u + v)\ (s + t)$; and

$u\ (s + t) + v\ (s + t)$ as $us + ut + vs + vt$; and

$a\ (b + c + d)$ as $ab + ac + ad$

2. Addition and Subtraction of Algebraic Expressions

Consider the expression $(5x + 3y + 6) + (2x + 5y + 2)$. Using the associative property, we can write it as:

$$5x + 3y + 6 + 2x + 5y + 2$$

Now, using the commutative property we can write it as:

$$5x + 2x + 3y + 5y + 6 + 2$$

Finally, using the distributive property it becomes:

$$(5 + 2)\,x + (3 + 5)\,y + 6 + 2$$

Therefore, $(5x + 3y + 6) + (2x + 5y + 2) = 7x + 8y + 8$.

Most of you could have written the sum of the above expressions upon inspection, and that is the preferred method. However, if asked to justify your answer you must be able to give the means by which it was reached.

Now consider the subtraction of two expressions: $(x + y - 2) - (3x + 5y + 6)$. In this case we must appeal to the results of elementary algebra and write: $(x + y - 2) - (3x + 5y + 6) = (x + y - 2) + (-3x - 5y - 6)$. We now have an addition and can see that the answer is $-2x - 4y - 8$.

Questions

Explain how the right-hand side of each of the expressions below is obtained by using a number property, definition, or result of elementary algebra. If any statement is not true, correct it.

1. $x^2 + xy = x(x + y)$
2. $(a + 3) \div b = (a + 3) \cdot \frac{1}{b}$
3. $u\,(s + t) \pm v\,(s + t) = u + v\,(s + t)$
4. $8x - (3x + 2) = 8x + (-3x + 2)$
5. $(r + s)\,(u + v) = r\,(u + v) + s\,(u + v)$
6. $3x^2 \cdot 7y = 21x^2y$
7. $(a + b) \cdot c = (a + c) \cdot (b + c)$

8. $(x \cdot y) \cdot (r \cdot s) \cdot (3 + s) = x \cdot y \cdot r \cdot s\,(3 + s)$

9. $3 - (7 - 2s) = 3 + (-7 - 2s)$

Rewrite each of the following expressions so that it does not contain parentheses or brackets.

10. $-7x - (y - 3)$

11. $(s - 3) - 3t$

12. $(x + y) \div 3$

13. $(a - b) - (a + b)$

14. $(a + b) \cdot (a + b)$

15. $2\,[(3x - 2y) - 4\,(x + y)]$

16. $3y - (2y + 3x - (2x + 3y))$

In each of the following, (a) find the sum of the expressions, and (b) subtract the second expression from the first.

17. $2a + b + c$ and $a + 2b - c$

18. $4x + 3y - 7$ and $2x - 5y - 2$

19. $3(s + t)$ and $-2(2s + t)$

20. $-(a - b + c)$ and $3\,(2a - 4b + 6)$

3. Multiplication of Algebraic Expressions

In Chapter 3, when calculating with powers of ten, we observed that $10^m \cdot 10^n = 10^{m+n}$ when _m_ and _n_ were integers. Clearly, we could have made the same arguments for any number _x_, that is $x^n \cdot x^m = x^{n+m}$ when _n_ and _m_ are integers. This property of exponents, together with the number properties of the preceding section, guides us in multiplication of algebraic expressions.

Consider the product of the two expressions $2s^2$ and $-3st^3$. Using the commutative and associative properties, we can write their product as:

$$(2) \cdot (-3) \cdot s^2 \cdot s \cdot t^3$$

Now, using multiplication and the above property of exponents we may rewrite the product as

$$-6s^3t^3$$

At first glance it seems as though we have no number property that can help us to multiply $(a^2 + 2)(3a^2 + 4a + 1)$. Remember, however, that $3a^2 + 4a + 1$ represents a number, call it A temporarily, so we have an expression of the form $(a^2 + 2)A$ and can apply the distributive property to get $a^2 \cdot A + 2 \cdot A$. Hence, the product $(a^2 + 2)(3a^2 + 4a + 1)$ can be written as

$$a^2 (3a^2 + 4a + 1) + 2(3a^2 + 4a + 1)$$

Another application of the distributive property permits us to write:

$$(a^2)(3a^2) + a^2(4a) + a^2(1) + (2)(3a^2) + (2)(4a) + (2)(1)$$

Simplifying each term of the last expression yields

$3a^4 + 4a^3 + a^2 + 6a^2 + 8a + 2$, or, combining terms,

$3a^4 + 4a^3 + 7a^2 + 8a + 2$

When multiplying long algebraic expressions it is sometimes convenient to use the long method of multiplication as shown below.

$$\begin{array}{ll}
 & 3a^2 + 4a + 1 \\
 & \underline{a^2 + 2 \qquad\quad} \\
(1) & 3a^4 + 4a^3 + a^2 \\
(2) & \underline{\qquad\qquad\quad 6a^2 + 8a + 2} \\
(3) & 3a^4 + 4a^3 + 7a^2 + 8a + 2
\end{array}$$

Rows (1) and (2) are obtained by multiplying the expression $3a^2 + 4a + 1$ by a^2 and then by 2, respectively. Row (2) is simply placed so that terms with the same exponent are arranged vertically so that the final sum, row (3), may be easily obtained. Notice that when multiplying with this arrangement we are using the same reasoning as before. That is, we are using the distributive property.

Questions

1. If m and n are positive integers, show that $(a^n)^m = a^{nm}$.
2. If n is a positive integer, show that $(ab)^n = a^n b^n$.

Perform the indicated multiplications.

3. $2x \cdot x^3 \cdot x^5$
4. $(s^2 t) \cdot t^3$
5. $(3r)^4$
6. $(a^2 b^3)^3$
7. $3y \cdot (x^2 + y)$
8. $(2x - 4)\ (3x + 4)$
9. $(s^2 - st + t^2)\ (s + t)$
10. $(x - y)\ (x + y)$
11. $(x - y)\ (x - y)$
12. $(m - 1)\ (m + 2)\ (m - 4)$
13. $(4x + 2y)\ (3x + y)$
14. $(m^3 - 2m^2 + m + 5)\ (m^2 + 3m - 4)$
15. $(4x - 2y)\ (4x + 2y)$
16. $(x - y)^3$

4. Some Special Products

There are three products which occur so frequently that they should be singled out for special attention. These three products are:

$$(x - y)\ (x + y) = x^2 - y^2$$

$$(x + y)^2 = (x + y)\ (x + y) = x^2 + y^2 + 2\,x\,y$$

$$(x - y)^2 = (x - y)\ (x - y) = x^2 + y^2 - 2\,x\,y$$

It should be understood that in these products, x and y may be any algebraic expressions. For example, if we replace x and y in the first expression by 25 and 3, respectively, we have

$$(25 - 3)\ (25 + 3) = 25^2 - 3^2$$

On the other hand, if we replace x and y in the first expression by $2S^2$ and t^3, respectively, we have

$$(2S^2 - t^3)\ (2S^2 + t^3) = (2S^2)^2 - (t^3)^2$$

To emphasize the fact the above special products involve arbitrary algebraic expressions, let us rewrite them using A and B to denote two arbitrary algebraic expressions.

(1) $(A - B)\ (A + B) = A^2 - B^2$

(2) $(A + B)\ (A + B) = A^2 + B^2 + 2AB$

(3) $(A - B)\ (A - B) = A^2 + B^2 - 2AB$

For example, to find the product of $(a^2 - b^2 + 1)$ and $(a^2 - b^2 - 1)$ we can think of $a^2 - b^2$ as A and of 1 as B, and have a product of the form

$$(A + B)\ (A - B)$$

Therefore, $(a^2 - b^2 + 1)\ (a^2 - b^2 - 1) = (a^2 - b^2)^2 - 1^2$.

We can expand this product further if we notice that $(a^2 - b^2)^2$ has the form of our third special product $(A - B)\ (A - B)$. Thus

$$(a^2 - b^2)^2 - 1^2 = (a^2)^2 + (b^2)^2 - 2\ (a^2)\ (b^2) - 1^2, \text{ or simply}$$

$$a^4 + b^4 - 2a^2b^2 - 1$$

Questions

1. Is $(A + B)\ (A - B)$ the same as $(A - B)\ (A + B)$? Explain.
2. Express in words, the identity $(A - B)\ (A + B) = A^2 - B^2$.
3. Express in words, the identity $(A + B)\ (A + B) = A^2 + B^2 + 2AB$ and $(A - B)\ (A - B) = A^2 + B^2 - 2AB$.
4. Work the example in this section $(a^2 - b^2 + 1)\ (a^2 - b^2 - 1)$ by the long method of multiplication.

Perform the indicated multiplications using the special products whenever possible.

5. $(3x + 4y)\ (3x - 4y)$

6. $(2r - t)^2$

7. $(\frac{1}{2} m + \frac{3}{4} v)^2$

8. $(x - 2)(x + 5)(x + 2)$

9. $(s + 2t + 3)(s + 2t - 3)$

10. $(x - 3y - z)^2$

11. $(m - 2v)(m + 2v)(m^2 + 4v^2)$

12. $[3(x + y) - 2][3(x + y) + 4]$

13. $(a + b - c - d)^2$

14. $(3x + y)^3$

15. $[(x + y)^2 - (x - y)^2][(x + y)^2 + (x - y)^2]$

5. Factoring

Very often, in the simplification of algebraic expressions, it is helpful to write a given algebraic expression as the product of other algebraic expressions, called its factors. There are a few basic steps to follow when attempting to factor an algebraic expression. Although these steps will not enable you to factor any given algebraic expression, they do provide a systematic procedure in many cases.

When all of the terms of an algebraic expression have a common factor we can use the distributive property. This procedure should always be tried first. For example:

$$2x^3 + 3x^2 + 6x = (2x^2 + 3x + 6) \cdot x$$

$$4(a + b) + (a - b)(a + b) = [4 + (a - b)] \cdot (a + b)$$

$$\begin{aligned} uw + vw + uy + vy &= (uw + vw) + (uy + vy) \\ &= (u + v) \cdot w + (u + v) \cdot y \\ &= (u + v) \cdot (w + y) \end{aligned}$$

Notice that in each of the above examples we have expressed the given algebraic expression as the product of other algebraic expressions (factors).

Whenever we see expressions of the form $A^2 - B^2$, $A^2 + B^2 + 2AB$, or $A^2 + B^2 - 2AB$, we should immediately associate them with the factors $(A - B)(A + B)$, $(A + B)^2$, and $(A - B)^2$ respectively. For example:

To factor $x^2 - 12x + 36$, notice that two of the terms in this expression are perfect squares; x^2 and $36 = 6^2$, and the third term is -2 times the product of x and 6. (i.e.: $x^2 - 12x + 36 = x^2 + 6^2 - 2(6)x$.) Thus we have an expression of the form $A^2 + B^2 - 2AB$ and it factors into $(x - 6)(x - 6)$.

If we factor $r^4 - 16$, we get

$$r^4 - 16 = (r^2)^2 - (4)^2$$

$$= (r^2 - 4)(r^2 + 4)$$

but, $r^2 - 4$ is also the difference of two squares and equals $(r - 2)(r + 2)$. Hence,

$$r^4 - 16 = (r - 2)(r + 2)(r^2 + 4)$$

Consider the expression $(x - 2)^2 + 14(x - 2) + 49$. It can be written as

$$(x - 2)^2 + (7)^2 + 2(7)(x - 2)$$

Thus, we have an expression of the form

$$A^2 + B^2 + 2AB \text{ where } A = (x - 2) \text{ and } B = 7.$$

So, $$(x - 2)^2 + 14(x - 2) + 49 = ((x - 2) + 7)((x - 2) + 7)$$

$$= (x + 5)^2$$

Questions

Factor the following expressions completely.

1. $3x - 18$
2. $3z^2 - 27$
3. $3x(2x + 5) + 4(2x + 5)$
4. $s^2 - 8s + 16$
5. $144a^8 - b^2$

6. $(a + b)^2 - (c + d)^2$

7. $x^2y - 2xy^2 + y^3$

8. $(x - 2)^2 + 4\,(x - 2)\,(y + 4) + 4\,(y + 4)^2$

6. Division of Polynomials in One Variable

It is often necessary to divide a polynomial expression in one variable by another in the same variable. For example, how do we divide $(2x^2 - 18x + 20)$ by $(x - 7)$? Before we attempt to divide polynomials, let us review a method for number division.

Suppose you were asked to divide 1760 by 49. The usual long division algorithm is familiar to most of us, but there is another way to approach the problem. We begin by making guesses. First, let's try 30. If we multiply 30 times 49, we get 1470, which we then subtract from 1760 (step 1). Notice that we have 290 left over, so we guess again, say 5, multiply 5 times 49 and subtract the result from 290 (step 2). Observe that we have taken 35 factors of 49 from 1760 and have a remainder of 45. We can summarize our results as

$$\begin{array}{rlc}
49\,\overline{)\,1760} & & \underline{\text{Choices}} \\
\left.\begin{array}{r}\underline{1470}\\ 290\end{array}\right\} & \text{step 1} & 30 \\
\left.\begin{array}{r}\underline{245-}\\ 45\end{array}\right\} & \text{step 2} & 5
\end{array}$$

$$1760 = 49 \cdot 35 + 45$$

$$\text{or}\quad \frac{1760}{49} = 35 + \frac{45}{49}$$

We can divide polynomials by this same "method of exhaustion." In fact, this procedure is probably easier for polynomials than it is for numbers. Consider the division of $(2x^2 - 18x + 20)$ by $(x - 7)$. If we wisely pick $2x$ as our first choice, then notice that we eliminate the first term of the polynomial when we multiply $(2x)\,(x - 7)$ and subtract it from $2x^2 - 18x + 20$ (step 1).

$$\begin{array}{rrc}
 & x-7\,\overline{)\,2x^2-18x+20} & \underline{\text{Choices}} \\
 & & 2x \\
\text{step 1} & \left\{\begin{array}{r}\underline{2x^2-14x}\\ -4x+20\end{array}\right. & \\
\text{step 2} & \left\{\begin{array}{r}\underline{-4x+28}\\ -8\end{array}\right. & -4
\end{array}$$

Next we choose -4 and repeat the process (step 2). We are left with a remainder of -8 and can summarize our results as

$$2x^2 - 18x + 20 = (2x - 4)\ (x - 7) + (-8)$$

or

$$\frac{2x^2 - 18x + 20}{2x - 4} = x - 7 + \frac{-8}{2x - 4}$$

Here is another example:

$$\begin{array}{lll}
 & & \underline{\text{Choices}} \\
x^2 - 2x + 1 & \overline{/x^5 \qquad\qquad\qquad\qquad\qquad} & x^3 \\
 & \left\{\begin{array}{l}\underline{x^5 - 2x^4 + x^3} \\ \qquad 2x^4 - x^3\end{array}\right. & 2x^2 \\
 & \qquad\quad \left\{\begin{array}{l}\underline{2x^4 - 4x^3 + 2x^2} \\ \qquad 3x^3 - 2x^2\end{array}\right. & 3x \\
 & \qquad\qquad\qquad \left\{\begin{array}{l}\underline{3x^3 - 6x^2 + 3x} \\ \qquad 4x^2 - 3x\end{array}\right. & 4 \\
 & \qquad\qquad\qquad\qquad\quad \left\{\begin{array}{l}\underline{4x^2 - 8x + 4} \\ \qquad 5x - 4\end{array}\right. &
\end{array}$$

After four steps we are left with a remainder of $5x - 4$ and can summarize our results as

$$x^5 = (x^3 + 2x^2 + 3x + 4)\ (x^2 - 2x + 1) + (5x - 4)$$

or

$$\frac{x^5}{x^2 - 2x + 1} = x^3 + 2x^2 + 3x + 4 + \frac{5x - 4}{x^2 - 2x + 1}$$

Questions

Divide:

1. $(t^2 - 7t + 10)$ by $(t - 5)$
2. $(y^3 - 4y^2 - 2 + 5y)$ by $(y - 1)$
3. $(6x^4 + 7x^3 + 12x^2 + 10x + 1)$ by $(2x^2 + x + 4)$
4. $(x^5 - 1)$ by $(x^2 + 1)$

5. $(x^5 + x^3 + x)$ by $(x + 1)$

6. $(6t^4 - 11t^3 - 12t^2 + 3t + 7)$ by $(2t - 1)$

7. Algebraic Fractions

An algebraic fraction is just the quotient of two algebraic expressions. To deal with algebraic fractions it is useful to recall certain properties of numbers.

The denominator of a fraction cannot equal zero. Therefore, when we write an algebraic fraction, say $\frac{x + 7}{x + 3}$, we must exclude any value of the variable which makes the denominator zero, in this case $x = -3$.

For each number s, there is a unique number which can be written as $\frac{1}{s}$ such that their product is one $(s \cdot \frac{1}{s} = 1)$. Such numbers are called multiplicative inverses of each other.

We can define division in terms of multiplication by multiplicative inverse:

$$a \div b = \frac{a}{b} = a \cdot \frac{1}{b}$$

and

$$\frac{1}{ab} = \frac{1}{a} \cdot \frac{1}{b}$$

These number properties must be kept in mind when working with algebraic fractions. Consider, for example,

$$\frac{(x - 3)\ (x - 2)}{(x - 3)\ (x - 1)}$$

Using number properties, we can rewrite this expression as

$(x - 3) \cdot \frac{1}{(x - 3)} \cdot \frac{(x - 2)}{(x - 1)}$ or since $x - 3$ and $\frac{1}{x - 3}$ are multiplicative inverses--that is, $(x - 3) \cdot \frac{1}{(x - 3)} = 1$, we can write the expression as

$$\frac{x - 2}{x - 1}$$

provided that x is not equal to 3 or to 1. When you have recorded the final result $\frac{x - 2}{x - 1}$, it is easy to forget the fact that in order to arrive at that result you assumed that x was not equal to 3.

8. Addition and Subtraction of Algebraic Fractions

Addition and subtraction of algebraic fractions with common denominators are straightforward operations obtained from number properties in the following way:

$$\frac{a}{b} + \frac{c}{b} = a \cdot \frac{1}{b} + c \cdot \frac{1}{b} = (a + c) \cdot \frac{1}{b} = \frac{a + c}{b}$$

$$\frac{a}{b} - \frac{c}{b} = a \cdot \frac{1}{b} - c \cdot \frac{1}{b} = (a - c) \cdot \frac{1}{b} = \frac{a - c}{b}$$

Addition and subtraction of fractions with different denominators are performed by rewriting the fractions so that they have a common denominator. Suppose we wish to add $\frac{a}{b} + \frac{c}{d}$. If we multiply the numerator and denominator of a fraction by the same number we do not change the value of the fraction--because we are just multiplying by 1. Thus we can write $\frac{a}{b} + \frac{c}{d}$ as $\frac{a}{b} \cdot \frac{d}{d} + \frac{c}{d} \cdot \frac{b}{b} = \frac{ad}{bd} + \frac{cb}{bd} = \frac{ad + cb}{bd}$.

Here is another example. We can add $\frac{1}{x + 3}$ and $\frac{x^2}{2x - 1}$ by multiplying these fractions by $\frac{2x - 1}{2x - 1}$ and $\frac{x + 3}{x + 3}$, respectively. We have $\frac{1}{x + 3} + \frac{x^2}{2x - 1} = \frac{1\ (2x - 1)}{(x + 3)\ (2x - 1)} + \frac{x^2\ (x + 3)}{(2x - 1)\ (x + 3)}$ and since we now have a common denominator, we can add to obtain

$$\frac{(2x - 1) + (x^3 + 3x^2)}{(x + 3)\ (2x - 1)} = \frac{x^3 + 3x^2 + 2x - 1}{2x^2 + 5x - 3}$$

In the expression $\frac{2s}{s^2 - 4} + \frac{s^2}{s^2 + 4s + 4}$ we notice that the denominators can be factored so that $\frac{2s}{(s^2 - 4)} + \frac{s^2}{s^2 + 4s + 4} = \frac{2s}{(s + 2)\ (s - 2)} + \frac{s^2}{(s + 2)\ (s + 2)}$. Multiplying the first expression by $\frac{s + 2}{s + 2}$ and the second by $\frac{s - 2}{s - 2}$ yields a common denominator,

$$\frac{2s}{(s + 2)\ (s - 2)} + \frac{s^2}{(s + 2)\ (s + 2)} = \frac{2s\ (s + 2)}{(s + 2)\ (s - 2)\ (s + 2)} + \frac{s^2\ (s - 2)}{(s + 2)(s + 2)\ (s - 2)}$$

Then adding,

$$\frac{2s\ (s + 2) + s^2\ (s - 2)}{(s + 2)\ (s + 2)\ (s - 2)} = \frac{2s^2 + 4s + s^3 - 2s^2}{(s + 2)^2\ (s - 2)} = \frac{4s + s^3}{(s + 2)^2\ (s - 2)}$$

Questions

Simplify, when possible, the following algebraic fractions.

1. $\frac{a^2 + b^2}{a^2 - b^2}$

2. $\frac{s^3 + 9s^2 + 20s}{s^2 + 9s + 20}$

3. $\frac{r^4 - s^4}{r^2 - s^2}$

4. $\frac{(x - 5)(x + 3)}{(x + 3)(x + 4)}$

The following simplifications are examples of common mistakes. Explain the faulty reasoning in each case.

5. $\frac{7x - 2}{7x} = -2$

6. $\frac{3u + 7}{3u + 8} = \frac{7}{8}$

7. $\frac{5r}{7} - \frac{2r}{6} = \frac{3r}{1}$

8. $\frac{x^2 - 2x + 5}{x^2 - 2x + 8} = \frac{5}{8}$

Carry out the indicated operations and simplify when possible:

9. $\frac{4}{x - 1} + \frac{3}{x - 2}$

10. $\frac{1}{3 - r} - \frac{1}{r}$

11. $\frac{a}{a + 1} + 2$

12. $y - \frac{y}{3}$

13. $\frac{3s}{s^2 + 4s + 3} - \frac{s}{s^2 - 9}$

9. Multiplication and Division of Algebraic Fractions

In order to multiply and divide fractions we must recall that for two fractions $\frac{a}{b}$ and $\frac{c}{d}$:

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$$

$\frac{a}{b} \div \frac{c}{d} = \frac{a}{b} \cdot \frac{d}{c}$. (To find the quotient of two fractions, invert the divisor and multiply.) To illustrate this latter property, consider the following proof:

$$\frac{a}{b} \div \frac{c}{d} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{\frac{a}{b} \cdot \frac{d}{c}}{\frac{c}{d} \cdot \frac{d}{c}} = \frac{\frac{a}{b} \cdot \frac{d}{c}}{\frac{cd}{cd}} = \frac{\frac{a}{b} \cdot \frac{d}{c}}{1} = \frac{a}{b} \cdot \frac{d}{c}$$

In the multiplication

$$\frac{3s^3}{4t^2} \cdot \frac{5t}{s^2} = \frac{15s^3t}{4t^2s^2}$$

notice that since the numerator and denominator have common factors, we may simplify the result by writing

$$\frac{15s^3t}{4t^2s^2} = \frac{15s\,(s^2t)}{4t\,(s^2t)} = \frac{15s}{4t} \cdot 1 = \frac{15s}{4t}$$

Dividing $(9u^3v^4 + 18u^4v^2 - 6uv)$ by $3u^2v^2$ is equivalent to multiplying

$$(9u^3v^4 + 18u^4v^2 - 6uv) \cdot \frac{1}{3u^2v^2}.$$

Applying the distributive property,

$$\frac{9u^3v^4}{3u^2v^2} + \frac{18u^4v^2}{3u^2v^2} - \frac{6uv}{3u^2v^2}$$

We can simplify by writing

$$\frac{9uv^2\,(u^2v^2)}{3\,(u^2v^2)} + \frac{18u^2\,(u^2v^2)}{3(u^2v^2)} - \frac{6(uv)}{3uv(uv)} = 3uv^2 + 6u^2 - \frac{2}{uv}$$

To multiply: $\frac{x - 2}{x^4 - 81} \cdot \frac{(x + 3)(x + 2)}{x^2 - 2x}$

It would be a waste of time to proceed by writing

$$\frac{x - 2}{x^4 - 81} \cdot \frac{(x + 3)(x + 2)}{x^2 - 2x} = \frac{(x - 2)(x^2 + 5x + 6)}{(x^4 - 81)(x^2 - 2x)} = \frac{x^3 + 3x^2 - 4x - 12}{x^6 - 2x^5 - 81x^2 + 162x}$$

because it is almost impossible to tell if the last expression can be simplified. A better method would be to see if any of the numerators or denominators can be factored before multiplying. In the case of this example, we can write

$$\frac{x - 2}{x^4 - 81} \cdot \frac{(x + 3)(x + 2)}{x^2 - 2x} = \frac{(x - 2)}{(x + 3)(x - 3)(x^2 + 9)} \cdot \frac{(x+2)(x+3)}{x(x - 2)}$$

$$= \frac{(x - 2)(x + 2)(x + 3)}{(x + 3)(x - 3)(x^2 + 9)(x)(x - 2)}$$

Now we can see that there are common factors in the numerator and the denominator of the product which can be written as:

$$\frac{(x + 2)}{(x - 3)(x^2 + 9)x}$$

In dividing $\frac{r^3 - r^2}{r + 3}$ by $(r - 1)$, our first step is to write the problem in terms of multiplication by the inverse.

$$\frac{r^3 - r^2}{r + 3} \div (r - 1) = \frac{r^3 - r^2}{r + 3} \cdot \frac{1}{r - 1}$$

Now we write:

$$\frac{r^3 - r^2}{r + 3} \cdot \frac{1}{r - 1} = \frac{r^2(r - 1)}{r + 3} \cdot \frac{1}{r - 1} = \frac{r^2(r - 1)}{(r + 3)(r - 1)} = \frac{r^2}{r + 3}$$

We sometimes encounter algebraic fractions in which the numerator and denominator are themselves composed of one or more fractions. Consider, for example, the expression:

$$\frac{1 + \frac{1}{a}}{1 - \frac{1}{a}}$$

Such expressions can be handled easily by first expressing the numerator and denominator as single fractions, thus obtaining a form you have already worked with, and then dividing fractions as usual. In this case we could write

$$\frac{1+\frac{1}{a}}{1-\frac{1}{a}} = \frac{\frac{a+1}{a}}{\frac{a-1}{a}} = \frac{a+1}{a} \cdot \frac{a}{a-1} = \frac{a+1}{a-1}$$

Questions

1. In Chapter 3, when calculating with powers of ten, we worked with expressions of the form $\frac{10^m}{10^n}$. Let us now consider expressions of the form $\frac{s^m}{s^n}$ where s is any positive number and m and n are positive integers. Using the fact that

$$\frac{s^m}{s^n} = \frac{\overbrace{s \cdot s \cdot s \cdot \ldots \cdot s}^{\underline{m} \text{ factors of } \underline{s}}}{\underbrace{s \cdot s \cdot s \cdot \ldots \cdot s}_{\underline{n} \text{ factors of } \underline{s}}}$$

Explain the following result:

$$\frac{s^m}{s^n} = \begin{cases} 1 \text{ if } m = n \\ s^{m-n} \text{ if } m > n \\ \frac{1}{s^{n-m}} \text{ if } m < n \end{cases}$$

Perform the indicated operations and simplify when possible.

2. $(3x^3y - 5xy^2 + 6x^3y^3) \div xy$

3. $(6u^3 - 9u^4v) \div 3uv$

4. $\frac{r^2 - 9}{r^2 + 2r} \cdot \frac{r+2}{r-3}$

5. $\frac{x^2 + x - 6}{x - 1} \div \frac{x-2}{3}$

6. $\frac{y - 3}{y} \div 2$

7. $\frac{4r + 8s}{3rs} \cdot \frac{9r^2 s}{3\,(r + s)}$

8. $\frac{yx - yz}{yx + yz} \cdot \frac{x}{x - z} \cdot \frac{x + z}{x}$

9. $\left[\frac{2t}{t - 1} + \frac{t^2}{t^2 - 1}\right] \div \frac{t^3}{1 - t}$

10. $\left[x + \frac{2}{x - 1}\right] \cdot \left[\frac{x^2}{x^2 - 1} \div \frac{x^3}{1 - x}\right]$

11. $\dfrac{2 + \frac{4}{5}}{\frac{2}{3} - 1}$

12. $\dfrac{\frac{1}{a} + \frac{1}{b}}{1 - \frac{a}{b}}$

13. $\dfrac{x}{1 - \frac{1}{x}}$

14. $\dfrac{s - \frac{s}{t}}{s + \frac{s}{t}}$

Determine whether or not the expressions in each of the following pairs are equivalent. If not, correct the expression on the right so that they are equivalent.

15. $\frac{xy - xz}{z}$; $x\,(\frac{y}{z} - \frac{z}{z})$

16. $\frac{c}{c - v}$; $\dfrac{1}{1 - \frac{v}{c}}$

17. $m\,(\frac{2n}{m + n})^2$; $n\left[\frac{4mn}{(m + n)^2}\right]$

18. $\frac{1}{r_o} - \frac{1}{r}$; $\frac{r_o - r}{r_o r}$

19. $\frac{m}{2}(r - \frac{n}{2})^2 + \frac{m}{2}(r + \frac{n}{2})^2$; $2mr^2 + \frac{1}{2}mn^2$

20. $3 - v$; $3(1 - \frac{v}{3})$

21. $\frac{u - v}{u + v}$; $\frac{1 - \frac{u}{v}}{1 + \frac{u}{v}}$

22. $\frac{p^2}{2m} + \frac{p^2}{2n}$; $2p^2 \frac{(m + n)}{2mn}$

23. $\frac{\frac{n}{m + n}}{\frac{m}{m + n}}$; $\frac{nm}{(n + m)^2}$

24. $\frac{\frac{7}{x} - \frac{7}{y}}{\frac{7}{y}}$; $\frac{y - x}{x}$

25. $(a + b)(1 - \frac{ab}{(a + b)a})$; a

26. $(\frac{1}{N})^{2N}$; $(\frac{1}{N})^{N + n} + (\frac{1}{N})^{N - n}$

Appendix 2: What Can We Do to Equations?

An equation is simply a statement that two expressions are equal. Thus,

$$2x + 5y = 6xy - 7,$$

$$x - a = 2y^2,$$

and

$$y = 3x + 4$$

are equations. Either side of an equation may have any number of terms. For example, the third equation above has one term on the left side (y) and two terms on the right side (3x and 4).

We can manipulate equations in many ways, depending on what we want to do. In a given equation we may wish to express one quantity in terms of the others, or solve for the unknown quantity, or isolate certain terms from others. Sometimes we have to work with two or more equations simultaneously. In all these cases it is necessary to know which manipulations are permitted, so as not to invalidate the original equality.

The purpose of this Appendix is to discuss some of the more common manipulations that are used when we work with equations.

1. Adding a Well-Chosen Zero

To manipulate one side only of an equation without invalidating it, we need to know two properties of numbers.

The first of these is

(a) Zero is the only number for which

$$x + 0 = x$$

for any number $\underline{x}$.

Property (a) is usually worded, "Zero added to any number does not change that number." Since an equation becomes an equality of numbers when a number is substituted for the variables, we can use any property of numbers to manipulate an equation into another form.

To illustrate the use of this property, sometimes called "adding a well-chosen zero," we consider the following equation:

$$y = x^2 + 6x + 2$$

and ask the question, "What is the smallest value that y can have in this equation?"

Since the expression contains an x^2 and an x term, we would like to combine them into a square of a sum. Recall that

$$(a + b)^2 = a^2 + 2ab + b^2$$

Here we have

$$x^2 + 6x \quad \text{or} \quad x^2 + (2)(3)x$$

To make this a perfect square we need to add 9. But we must also subtract 9 to keep the same value of y. Our "0" = 9 - 9.

$$y = x^2 + 6x + (9 - 9) + 2$$

Recognizing $x^2 + 6x + 9$ as $(x+3)^2$, we can write

$$y = (x + 3)^2 - 7$$

The $(x+3)^2$ term is ≥ 0, hence, its smallest value is zero. The smallest value of y is then -7.

In general, to make a "perfect square" from the expression

$$x^2 + mx + n$$

we have to add $0 = (\frac{m}{2})^2 - (\frac{m}{2})^2$

$$x^2 + 2\,(\frac{m}{2})x + (\frac{m}{2})^2 + n - (\frac{m}{2})^2$$

$$= (x + \frac{m}{2})^2 + \frac{4n - m^2}{4}$$

Here is another example of adding "a well-chosen zero." Given the equation

$$y = \frac{x^2 + x + 1}{x^2 + 3x + 4}$$

suppose we are required to divide until the degree of the numerator is smaller than the degree of the denominator. It is much easier to add 0 to the numerator chosen in such a way as to make the original numerator equal to the denominator. We see here that the numerator needs the quantity 2x + 3 added to it to make this so. Thus, we have

$$y = \frac{x^2 + x + 1}{x^2 + 3x + 4} = \frac{x^2 + x + 1 + [(2x + 3) - (2x + 3)]}{x^2 + 3x + 4}$$

$$= \frac{x^2 + 3x + 4 - (2x + 3)}{x^2 + 3x + 4} = 1 - \frac{2x + 3}{x^2 + 3x + 4}$$

Since we are working with one side only of an equation, we are actually working with an expression. As we have just shown, zero may be added to any expression without changing its value. It is common to rewrite an expression like the following:

$$\frac{1}{1 + t}$$

Adding "a well-chosen zero" would result in

$$\frac{1}{1 + t} = \frac{1 + (t - t)}{1 + t} = \frac{1 + t}{1 + t} - \frac{t}{1 + t}$$

$$= 1 - t\left(\frac{1}{1 + t}\right)$$

We can repeat this process of adding $0 = t - t$ to the numerator inside the parentheses.

$$y = \frac{1}{1 + t} = 1 - t\left(\frac{1 + t - t}{1 + t}\right) = 1 - t\left(1 - t\frac{1}{1 + t}\right)$$

$$= 1 - t + t^2\left(\frac{1}{1 + t}\right) = 1 - t + t^2\left(\frac{1 + t - t}{1 + t}\right)$$

$$= 1 - t + t^2 - t^3\left(\frac{1}{1 + t}\right)$$

This expansion may be carried on to any number of terms. As you would expect, it gives the same result as does ordinary long division.

Incidentally, the example that we have chosen also illustrates the expansion of a power series. Notice that each term contains a higher power of t than the one preceding it. We speak of the term not containing t as the "zero-order term," the term containing t to the first power as the "first-order term," etc. For t very much less than 1, each term is significantly less than the one preceding it. When t is a physical number, it frequently suffices to retain only the first-order term.

$$\frac{1}{1+t} \approx 1 - t$$

This approximation is valid for values of t so small that t^2 is about the same value as the error in the value of t itself.

As another example of the expansion of a power series, we shall rewrite the expression

$$\frac{1}{1-x^2}$$

as follows:

$$\begin{aligned}\frac{1}{1-x^2} &= \frac{1+(x^2-x^2)}{1-x^2} = \frac{1-x^2}{1-x^2} + \frac{x^2}{1-x^2}\\ &= 1 + x^2\left(\frac{1}{1-x^2}\right) = 1 + x^2\left(\frac{1+x^2-x^2}{1-x^2}\right)\\ &= 1 + x^2 + x^4\left(\frac{1}{1-x^2}\right) = 1 + x^2 + x^4\left(\frac{1+x^2-x^2}{1-x^2}\right)\\ &= 1 + x^2 + x^4 + x^6\left(\frac{1}{1-x^2}\right)\end{aligned}$$

This expansion contains only even order terms. It has no 1st, 3rd, or 5th order terms. However, if we rewrite the expansion as

$$1 + (x^2) + (x^2)^2 + (x^2)^3\left(\frac{1}{1-x^2}\right)$$

then we can, for example, speak of the (x^2) term as the first order term in x^2, or $(x^2)^2$ as the second order term in x^2, etc.

Questions

1. Express $y = x^2 + 6x + 11$ as the square of the sum of x and a number, plus a constant.
2. Complete the square of the following quadratics:
   (a) $x^2 + 2x - 1 = 0$
   (b) $3x^2 - 2x + 6 = 0$
   (c) $5x^2 - 7x + 16 = 0$
3. Expand $\frac{1}{1-x^3}$ to second order in x^3 by adding well-chosen zeros.

4. Expand $\frac{2}{1 - \frac{x^2}{2}}$ to second order in x, and determine the accuracy of the approximation when $x = 0.1$.

2. Multiplying by a Well-Chosen One

The second property of numbers that we can also apply to expressions is

(b) One is the only number for which

$$x \cdot 1 = x$$

for any number x.

Property (b) is worded, "One multiplied by any number does not change that number."

As an example of modifying an expression using this property, sometimes called "multiplying by a well-chosen 1," suppose we want to find an approximate value of

$$\frac{\sqrt{x+h} - \sqrt{x}}{h}$$

where $h << x$ (very much less than). It is not much help to set $h = 0$ here, because then the expression reduces to zero divided by zero, which is meaningless. However, by multiplying by "a well-chosen 1,"

$$\frac{\sqrt{x+h} - \sqrt{x}}{h} = \frac{(\sqrt{x+h} - \sqrt{x})}{h} \frac{(\sqrt{x+h} + \sqrt{x})}{(\sqrt{x+h} + \sqrt{x}}$$

$$= \frac{x + h - x}{h(\sqrt{x+h} + \sqrt{x})} = \frac{1}{\sqrt{x+h} + \sqrt{x}}$$

Now, for $h << x$, we can approximate this expression by setting h in the denominator equal to zero.

Thus, $\frac{\sqrt{x+h} - \sqrt{x}}{h} \approx \frac{1}{\sqrt{x+0} + \sqrt{x}} = \frac{1}{2\sqrt{x}}$.

Multiplying by "a well-chosen 1" is useful in factoring expressions. Consider the expression

$$u + v$$

If we were asked to factor out a u, we might say that this is not possible since no factors of u are in the second term. But if we multiply by the "well-chosen 1" where $1 = u \cdot \frac{1}{u}$ and move the $\frac{1}{u}$ inside, we have

$$u + v = u \cdot \frac{1}{u}(u + v) = u\left(\frac{u}{u} + \frac{v}{u}\right)$$
$$= u\left(1 + \frac{v}{u}\right)$$

This particular example arises quite often when v and u are physical numbers and $v << u$. Factoring in this way enables us to see the contribution of v as a fraction of u.

Another example of manipulating an expression to see more easily the contribution of each term to the value of the expression is the following:

$$y = a_3x^3 + a_2x^2 + a_1x + a_o$$

For large x each term becomes successively smaller, and by introducing a well-chosen 1,

$$y = \frac{a_3x^3}{a_3x^3}(a_3x^3 + a_2x^2 + a_1x + a_o)$$
$$= a_3x^3\left(1 + \frac{a_2}{a_3} \cdot \frac{1}{x} + \frac{a_1}{a_3} \cdot \frac{1}{x^2} + \frac{a_o}{a_3} \cdot \frac{1}{x^3}\right)$$

we see how much smaller than 1 the $\frac{a_2}{a_3}$ term is.

Questions

1. Show that for $h << 1$

$$\sqrt{1 + h} \approx 1 + \frac{h}{2}$$

[Hint: Let $\sqrt{1 + h} = 1 + (-1 + \sqrt{1 + h})$, then multiply this expression by $\frac{(\sqrt{1 + h} + 1)}{(\sqrt{1 + h} + 1)}$.]

2. Approximate to three places using the ideas of this section:

   (a) $\frac{1}{0.95}$

   (b) $\frac{1}{1.01}$

3. What Can Be Done to an Equation by Working With Both Sides Without Invalidating the Equation?

We state four properties of numbers which, as we have cited earlier, are applicable to equations since equations reduce to an equality of numbers when numbers are substituted for the variables.

(a) If $\underline{a}$, $\underline{b}$, $\underline{c}$ are numbers and $a = b$, then $a + c = b + c$.

This is usually verbalized, "One can add the same number to both sides of an equality without changing the equality."

(b) If $\underline{a}$, $\underline{b}$, $\underline{c}$ are numbers and $a = b$, then $a \cdot c = b \cdot c$.

In words, "One can multiply both sides of an equality by the same number." We point out, however, that multiplying by zero is useless, since this reduces all equations to the identity $0 \equiv 0$.

(c) If $\underline{a}$, $\underline{b}$, $\underline{c}$ are numbers, $a = b$, and $c \neq 0$, then $\frac{a}{c} = \frac{b}{c}$.

That is, "One can divide both sides of an equality by the same nonzero number."

(d) If $a = b$, then $a^n = b^n$, and in particular, $\frac{1}{a} = \frac{1}{b}$.

In words, "One can raise both sides of an equality to the same power."

We illustrate each of (a), (b), and (c) above by solving

$$r = \frac{t}{t + 1}$$

for $\underline{t}$ in terms of $\underline{r}$.

The general approach to this type of equation is to clear the equation of fractions and then isolate the unknown on one side of the equation.

We start with property (b), that is, multiply both sides of the equation by (t + 1).

$$r(t + 1) = \frac{t}{(t + 1)} \cdot (t + 1)$$

Canceling the (t + 1) factors on the right side we have

$$r(t + 1) = t$$

We then distribute the product on the left over the sum.

$$rt + r = t$$

To isolate t on one side, we add -rt to both sides of this equation (property (a))

$$-rt + rt + r = t - rt$$

which becomes

$$r = t - rt$$

Using the distributive law again, we factor t out of the right side.

$$r = t(1 - r)$$

Then dividing both sides by 1 - r (property (c)),

$$t = \frac{r}{1 - r}$$

Notice that when we used property (c), we divided by (1 - r). This requires that $r \neq 1$, because otherwise $1 - r = 0$. In general, when dividing by polynomials that contain a variable, we must be sure that the variable does not have a value that makes the polynomial zero.

We shall illustrate property (d) with the following equation

$$p = \frac{mv}{\sqrt{1 - \frac{v^2}{c^2}}}$$

by solving the equation for v, that is, getting v all by itself on one side of the equation. Before we start, it would be helpful to note some of the restrictions that must be placed on the values of the variables in this equation. Clearly, $c \neq v$, since $c = v$ would result in a zero denominator. Also $p \neq 0$ and $m \neq 0$; otherwise the equation reduces to $0 = 0$, which is not very useful. These three restrictions, $c \neq v$, $p \neq 0$, and $m \neq 0$, will become more obvious as we proceed with the solution.

Let's do the work now. Squaring (property (d)) gives

$$p^2 = \frac{m^2v^2}{1 - \frac{v^2}{c^2}}$$

Next, multiply by $1 - \frac{v^2}{c^2}$. (This is not zero because $c \neq v$.)

$$p^2 (1 - \frac{v^2}{c^2}) = m^2v^2$$

Using the distributive law on the left yields

$$p^2 - \frac{p^2v^2}{c^2} = m^2v^2$$

To collect the v^2 terms we add $\frac{p^2v^2}{c^2}$ to both sides

$$p^2 - \frac{p^2v^2}{c^2} + \frac{p^2v^2}{c^2} = m^2v^2 + \frac{p^2v^2}{c^2}$$

Using the distributive law on the right we have

$$p^2 = v^2 (m^2 + \frac{p^2}{c^2})$$

To isolate v^2, we divide by $m^2 + \frac{p^2}{c^2}$. (This is not zero because $m \neq 0$, $p \neq 0$.)

$$\frac{p^2}{m^2 + \frac{p^2}{c^2}} = v^2$$

The final step is to take square roots.

$$v = \frac{p}{\sqrt{m^2 + \frac{p^2}{c^2}}}$$

Questions

Before manipulating, plan your steps. These problems will actually occur if you study physics.

1. $\frac{1}{2} mv^2 + \frac{1}{2}kx^2 = E$. Solve for v.
2. $T = 2\pi\sqrt{\frac{m}{k}}$. Solve for k.
3. Let $\frac{P_1}{P_2} = (\frac{v_1}{v_2})^N$. Solve for v_2.

4. $\frac{1}{a} + \frac{1}{b} = \frac{1}{c}$. Discuss the best way to rearrange this to calculate a from a given pair of values for b and c.

4. Solving the Quadratic Equation

As we have seen in Section 1 of this Appendix, when we add the "well-chosen" zero $(\frac{m}{2})^2 - (\frac{m}{2})^2$ to the quadratic equation

$$x^2 + mx + n = 0$$

the equation becomes $(x + \frac{m}{2})^2 + \frac{4n - m^2}{4} = 0$.

The more general quadratic

$$ax^2 + bx + c = 0$$

can be solved in the same manner.

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0$$

$$x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} = \frac{b^2}{4a^2} - \frac{c}{a}$$

$$(x + \frac{b}{2a})^2 = \frac{b^2 - 4ac}{4a}$$

$$x + \frac{b}{2a} = \pm\sqrt{\frac{b^2 - 4ac}{4a}}$$

$$x = -\frac{b}{2a} \pm \frac{1}{2a}\sqrt{b^2 - 4ac}$$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

This result is called the quadratic formula where a and b are the coefficients of the x^2 and x term respectively, and c is the value of the constant term.

$b^2 - 4ac$ in this formula is called the discriminant because it identifies the character of the roots of the quadratic. When the discriminant is zero, the roots are equal (they are $-\frac{b}{2a}$). When $b^2 - 4ac$ is greater than zero, the original equation has two solutions. When the discriminant is negative, there are no solutions because no number on the number line is the square root of a negative number.

Questions

1. Using the quadratic formula, show that the
   (a) sum of the roots $x_1 + x_2 = -\frac{b}{a}$.
   (b) product of the roots $x_1 x_2 = \frac{c}{a}$.
2. Which of the following equations (a) have equal roots, (b) are factorable, (c) have no solution?
   (a) $9 = x^2 + 6x$
   (b) $2x^2 - 7x + 10 = 0$
   (c) $12x^2 - 95x - 8 = 0$
   (d) $10x^2 - 41x - 156 = 0$
   (e) $4x^2 - 12x + 9 = 0$
   (f) $5x^2 - 3x + 2 = 0$
3. For what value of k will the roots of the following equations be equal?
   (a) $3x^2 + 4k = 5x$
   (b) $4(x - 1)^2 = 2 + kx$
   (c) $kx^2 - 3 + 2kx = 0$
4. Find the value of k if the product of the roots of $3x^2 - 2x - k = 0$ is 2.
5. Find the value of k if the roots of $x^2 + kx + 4$ differ by 3.

5. Substitution

We can substitute for any variable in an equation an expression that is equal to that variable. For example, consider the equation

$$y = ax^3 + bx^2 + cx + d$$

If, in addition to this,

$$x = u + v$$

then we are permitted to substitute (u + v) for x in the general equation wherever an x occurs. Thus, we write

$$y = a(u + v)^3 + b(u + v)^2 + c(u + v) + d$$

To take another example, suppose we had the following system of equations:

$$\begin{cases} s = v_1 t + \frac{1}{2}at^2 \\ F = m\,\frac{(v_2 - v_1)}{t} \\ v_2 - v_1 = at \end{cases}$$

and

$$t = \frac{2s}{v_1 + v_2}$$

Wherever we see a t in the system of equations, we can replace it with the equivalent expression $\frac{2s}{v_1 + v_2}$.

In this case, the three equations become

$$s = v_1 \left(\frac{2s}{v_1 + v_2}\right) + \frac{1}{2}a \left(\frac{2s}{v_1 + v_2}\right)^2$$

$$F = m\,\frac{(v_2 - v_1)}{\left(\frac{2s}{v_1 + v_2}\right)}$$

$$v_2 - v_1 = a\left(\frac{2s}{v_1 + v_2}\right)$$

Questions

1. Let it be given that $V = \frac{2\pi R}{T}$ and $a = \frac{2\pi V}{T}$. Express a in terms of R and T.
2. The following occur in elementary orbital problems:

$$a = \frac{4\pi^2 R}{T^2};\ F = ma;\ \frac{R^3}{T^2} = k$$

Solve for F in terms of k, m, and R.

3. $S = 2\pi r^2 + 2\pi rh$. Suppose r and h are related by $2r = h$. Find S in terms of r.
4. Let $y = \frac{1}{x}$ and $x = t + 2$. Express y in terms of t.
5. Let $p = \frac{P}{RT}$; $p = \frac{M}{V}$. Eliminate p and express P in terms of M, V, R, and T.

6. Express $\underline{v}$ in terms of $\underline{g}$ and $\underline{s}$ in the following:

$$t = \sqrt{\frac{2s}{g}}$$

$$v = gt$$

7. (a) From $E = \dfrac{mc^2}{\sqrt{1 - \dfrac{v^2}{c^2}}}$ and $p = \dfrac{mv}{\sqrt{1 - \dfrac{v^2}{c^2}}}$ eliminate $\underline{v}$ and thus express $\underline{E}$ in terms of $\underline{p}$, $\underline{m}$, and $\underline{c}$.

(b) For $p \ll mc$ expand your result to second order in $\dfrac{p}{mc}$.

6. What Can Be Done to Two Or More Equations?

When working with two or more equations, we can use any of the foregoing ideas on any member of the set of equations, namely, adding the same expression to both sides and multiplying or dividing both sides by the same (non-zero) expression. We can also substitute for a variable in an equation any expression that is equal to that variable. There are, in addition, the ideas of adding equations, multiplying equations, and dividing equations. We indicate these operations schematically by writing the following:

If A, B, C, D are expressions and if $A = B$ and $C = D$,

$$\text{then} \quad A + C = B + D$$

$$A \cdot C = B \cdot D$$

$$\frac{A}{C} = \frac{B}{D}.$$

Let us begin with the general solution of two equations in two unknowns.

$$a_1x + b_1y = c_1 \qquad (1)$$

$$a_2x + b_2y = c_2 \qquad (2)$$

Our plan is to eliminate one of the unknowns from this set, arriving at one equation in one unknown. We can do this by multiplying one of the equations by an appropriate constant such that when the two equations are then added, one of the unknowns drops out.

Specifically, if we multiply the first equation by $-\frac{a_2}{a_1}$, then add this result to the second equation, we get

$$\left(-\frac{a_2}{a_1}\right) a_1x + \left(-\frac{a_2}{a_1}\right) b_1y = \left(-\frac{a_2}{a_1}\right) c_1 \qquad \left(-\frac{a_2}{a_1}\right) \cdot (1)$$

$$a_2x + b_2y = c_2 \qquad (2)$$

$$0 + \left(b_2 - \frac{a_2}{a_1} b_1\right) y = c_2 - \frac{a_2}{a_1} c_1 \qquad \left(-\frac{a_2}{a_1}\right) \cdot (1) + (2)$$

where the notation at right indicates the operations being performed. We now solve this equation for y:

$$y = \frac{c_2 - \frac{a_2}{a_1} c_1}{b_2 - \frac{a_2}{a_1} b_1}$$

$$y = \frac{a_1c_2 - a_2c_1}{a_1b_2 - a_2b_1}$$

To find x, we take this value of y and substitute it in the first equation in place of y.

$$a_1x + b_1y = c_1$$

$$a_1x + b_1\left(\frac{a_1c_2 - a_2c_1}{a_1b_2 - a_2b_1}\right) = c_1$$

This equation reduces to

$$x = \frac{b_2c_1 - b_1c_2}{a_1b_2 - a_2b_1}$$

In a set of n equations in n unknowns one uses this same procedure to eliminate all the x's below the first and thus reduces the system to n - 1 equations in n - 1 unknowns. We illustrate by reducing a three-equation system to a two-equation system.

$$a_1x + b_1y + c_1z = d_1 \qquad (1)$$

$$a_2x + b_2y + c_2z = d_2 \qquad (2)$$

$$a_3x + b_3y + c_3z = d_3 \qquad (3)$$

We replace equation (2) by the sum of equation (2) and $(-\frac{a_2}{a_1})$ times equation (1), and we replace equation (3) by the sum of equation (3) and $(-\frac{a_3}{a_1})$ times equation (1).

$a_1 x + b_1 y + c_1 z = d_1$ (1)

$(b_2 - \frac{a_2}{a_1} b_1) y + (c_2 - \frac{a_2}{a_1} c_1) z = d_2 - \frac{a_2}{a_1} d_1$ $(2') = (-\frac{a_2}{a_1})$ (1) + (2)

$(b_3 - \frac{a_3}{a_1} b_1) y + (c_3 - \frac{a_3}{a_1} c_1) z = d_3 - \frac{a_3}{a_1} d_1$ $(3') = (-\frac{a_3}{a_1})$ (1) + (3)

Equations (2') and (3') contain two unknowns and these are solved as before for the two-equation case. $\underline{x}$ is then found by substituting the values of $\underline{y}$ and $\underline{z}$ into equation (1).

As an example, we include the solution of a three-equation system.

$x + \frac{1}{2} y + \frac{1}{2} z = 1$ (1)

$3x + 3y + 4z = 2$ (2)

$5x + 4y + z = 1$ (3)

$x + \frac{1}{2} y + \frac{1}{2} z = 1$ (1)

$0x + \frac{3}{2} y + \frac{5}{2} z = -1$ (2') = -3 (1) + (2)

$0x + \frac{3}{2} y - \frac{3}{2} z = -4$ (3') = -5 (1) + (3)

$x + \frac{1}{2} y + \frac{1}{2} z = 1$ (1)

$\frac{3}{2} y + \frac{5}{2} z = -1$ (2')

$0y + \frac{8}{2} z = 3$ (3") = (2') - (3')

$$\boxed{z = \frac{3}{4}}$$

$\frac{3}{2} y + \frac{5}{2} z = -1$ (2')

$\frac{3}{2} y + \frac{5}{2} (\frac{3}{4}) = -1$

$y = \frac{2}{3} (-\frac{15}{8} - \frac{8}{8})$

$$\boxed{y = -\frac{23}{12}}$$

$$x + \frac{1}{2}y + \frac{1}{2}z = 1 \qquad (1)$$

$$x + \frac{1}{2}(-\frac{23}{12}) + \frac{1}{2}(\frac{3}{4}) = 1$$

$$x = \frac{23}{24} - \frac{3}{8} + 1$$

$$\boxed{x = \frac{19}{12}}$$

Checks are most important. You should actually substitute these values for x, y, and z into equations (1), (2), and (3) to show that the equations are satisfied.

Occasionally it is desirable to divide one equation by another. We illustrate with the following:

$$\text{Let } x = r\theta$$
$$y = r\sqrt{1 - \theta^2}$$

If we want to solve for θ and r (in terms of x and y), we can eliminate r by dividing the first equation by the second and equating the quotients.

$$\frac{x}{y} = \frac{\theta}{\sqrt{1 - \theta^2}}$$

We can now solve for θ. Square both sides and multiply by $1 - \theta^2$

$$(\frac{x}{y})^2 = \frac{\theta^2}{1 - \theta^2}$$

$$(\frac{x}{y})^2 (1 - \theta^2) = \theta^2$$

$$\theta^2 (1 + (\frac{x}{y})^2) = (\frac{x}{y})^2$$

$$\theta^2 = \frac{(\frac{x}{y})^2}{1 + (\frac{x}{y})^2}$$

Taking the square root of each side:

$$\theta = \frac{\frac{x}{y}}{\sqrt{1 + (\frac{x}{y})^2}}$$

$$\theta = \frac{\frac{x}{y}}{\sqrt{\frac{y^2 + x^2}{y^2}}} = \frac{\frac{x}{y}}{\frac{1}{y}\sqrt{x^2 + y^2}}$$

$$= \frac{x}{\sqrt{x^2 + y^2}}$$

To find r, substitute the known value of θ into the first equation.

$$x = r\theta$$

$$x = r\,\frac{x}{\sqrt{x^2 + y^2}}$$

$$r = \sqrt{x^2 + y^2}$$

Thus
$$\begin{cases} r = \sqrt{x^2 + y^2} \\ \theta = \dfrac{x}{\sqrt{x^2 + y^2}} \end{cases}$$

Questions

1. Complete the algebra in the text discussion of the general solution of two equations in two unknowns and show that

$$x = \frac{b_2c_1 - b_1c_2}{a_1b_2 - a_2b_1}$$

2. If the following have solutions, solve. How many solutions are there?

(a) $x + y + 2z = 1$
$2x + y + 3z = 2$
$x + 2z = 1$

(b) $x + y + 2z = 1$
$2x + y + 3z = 2$
$x + z = 1$

(c) $x + y + 2z = 1$

$2x + 2y + 3z = 0$

$x + y + z = 0$

3. Suppose $A = \sqrt{xy}$ and $B = \sqrt{\frac{x}{y}}$. Discuss how to find x and y in terms of A and B alone.

4. Let $u = \frac{x}{x^2 - y^2}$

$v = \frac{-y}{x^2 + y^2}$

Solve for x and y.

[Hint: $u^2 + v^2 = \frac{x^2}{(x^2 - y^2)^2} + \frac{y^2}{(x^2 + y^2)^2} = \frac{1}{x^2 + y^2}$.]

7. Graphical Solution of Two Equations

Simultaneous equations in two unknowns may also be solved graphically. Since the intersection of the graphs of each equation is a point common to both graphs, this point must satisfy the equation of each graph, that is, it is a solution to the equations. For example, the two equations

$$5x - 2y = 4$$
$$4x + 3y = 17$$

each represent a straight line whose graph is shown in Fig. 1. The graphs intersect at the point (2,3), thus, the solution to the equations is $x = 2$ and $y = 3$.

Sometimes, as in the case of the two equations

$$2y - x = 8$$
$$2y - x = -3$$

their graphs do not intersect as shown in Fig. 2. The lines are parallel, the slopes are equal, and there is no common solution. In these cases, we call the set of equations inconsistent.

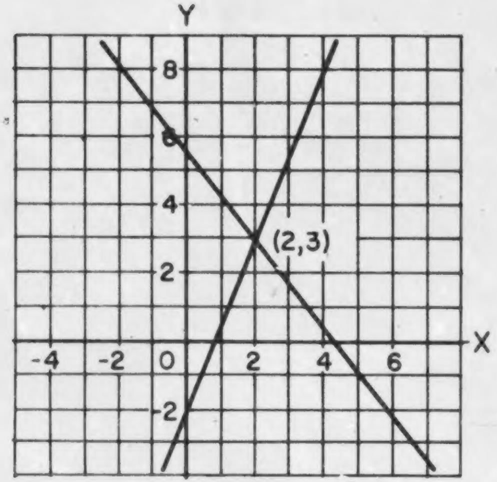


Fig. 1

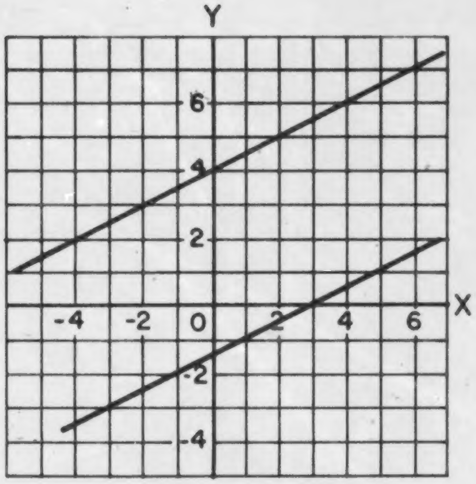


Fig. 2

Questions

1. Solve, when possible, the following sets of equations graphically:

(a) $2x - y = 5$
$3x + 2y = -7$

(b) $4x + 5y = 3$
$3x - 2y = 5$

(c) $5x - 3y = 4$
$10x - 6y = -1$

(d) $y - 2 = \frac{2}{3}(x + 1)$
$y + 1 = 6x$

(e) $x = \frac{1}{6}(1 + 4y)$
$3x - 2y = 4$

(f) $2x + 1 = 3(y - 4)$
$y = 5x$

2. Solve the set of equations in Fig. 1 and show that the solution is consistent with the graphical solution.

3. Find a graphical solution of the set of equations

$$x + y = 7$$
$$y = (x - 4)^2$$

4. Find the roots of the equation

$$2x^2 - 5x - 12 = 0$$

graphically, and then check your result by using the quadratic formula.

5. Some Precautions

We have indicated throughout this Appendix that manipulations performed on equations must be done so as not to invalidate the equation. In particular, we cited the cautions to be observed in multiplying or dividing by zero.

Let us look at these two restrictions more closely. Given the equation

$$x - 3 = 2$$

we shall multiply both sides by the quantity $(x - 2)$, getting

$$(x - 3)(x - 2) = 2(x - 2)$$

which becomes

$$x^2 - 5x + 6 = 2x - 4$$

$$x^2 - 7x + 10 = 0$$

$$(x - 5)(x - 2) = 0$$

This last equation has two solutions, $x = 5$ and $x = 2$. The original equation, however, has only one solution, $x = 5$. We see, then, that the equation we started with and the equation we ended up with are not equivalent. It should be apparent that when $x = 2$, the value of the multiplier we used is zero, which, in turn, led to the extra solution, $x = 2$. In situations like this, we call such roots extraneous.

Now we consider an example of dividing an equation by an expression containing a variable. If we have

$$(x - 3)(x - 2) = 4(x - 3)$$

and divide each side by the quantity $(x - 3)$, we get

$$\frac{(x - 3)}{(x - 3)}(x - 2) = 4\frac{(x - 3)}{(x - 3)}$$

$$x - 2 = 4$$

$$x = 6$$

Again, the first and last equations are not equivalent, but now the first equation has two solutions ($x = 6$ and $x = 3$) and the last equation has only one. This is because when $x = 3$ the value of the divisor is zero. So in this case we lose solutions.

In summary, multiplication or division of an equation by an expression containing a variable is prohibited for that value of the variable which reduces the expression to zero.

In Section 2 we made the following statement:

$$\text{If } a = b, \text{ then } a^n = b^n.$$

That is, both sides of an equality can be raised to the same power. Let us examine this idea further.

If we have

$$\sqrt{x + 1} = \sqrt{y}$$

and square both sides, we get

$$x + 1 = y$$

By this process, we do not lose any solutions because any pair of values for x and y that satisfy the first equation will satisfy the second equation. However, $x = -2$ and $y = -1$ satisfy the second equation, but these two values reduce the first equation to the statement $\sqrt{-1} = \sqrt{-1}$. This result makes no sense in the context of the number line because no number on the line is the square root of a negative number. Squaring the equation $\sqrt{x + 1} = \sqrt{y}$, then, has led to extraneous roots.

Another example of where squaring an equation leads to extraneous roots is the following. If

$$x = a$$

then $$x^2 = a^2$$

and $$x^2 - a^2 = 0$$

or $$(x - a)(x + a) = 0$$

which has the two solutions $x = a$, $x = -a$.

In general, when we raise both sides of an equation to an even power we will always pick up extra solutions. This is not to say that we are never allowed to raise an equation to an even power. It is just that when we perform this particular manipulation we must be aware of the consequences.

If we consider odd powers, no problem arises. If

$$x = a$$

then

$$x^3 = a^3$$

and

$$x = \sqrt[3]{a^3} = a$$

We neither gain nor lose solutions.

Questions

1. Solve for x and check for extraneous roots

$$\sqrt{x + 6} = x$$

2. Given the two equations

$$y = \sqrt{1 - x^2}$$

$$x = 4 + t^2$$

(a) What limitations are placed on x in the first equation?

(b) What is the smallest numerical value that x can have in the second equation?

(c) Can you substitute the expression for x from the second equation into the first equation? Explain.

3. Starting with the false equation $7 = 9$, subtract 8 from both sides.

$$7 - 8 = 9 - 8$$

$$\text{or} \quad -1 = +1$$

Then square both sides

$$(-1)^2 = (1)^2$$

which results in a true equation

$$1 = 1$$

Can you explain what has happened?

Appendix 3: Inequalities

1. Notation

An inequality is a statement that two quantities are not equal. If, for example, $a \neq b$, then either a is greater than b ($a > b$); or a is less than b ($a < b$). The symbols ">" and "<" denote the sense of the inequality. Remember that the tip of the inequality sign points toward the smaller quantity.

The "continued" inequality $a < b < c$

means $a < b$ and $b < c$

The statement $1 < x < 2$

means "x is between 1 and 2."

We never write $2 < x < -2$

for this means $2 < x$ and $x < -2$

which is not true for any x.

Instead, we would write $x > 2$ or $x < -2$

$a \geq b$ means a is equal to or greater than b.

$a \leq x \leq b$ is read as, "x is equal to or greater than a and equal to or less than b."

Finally, when we write

$a > 0$, we speak of a being positive

$a < 0$, we speak of a being negative

$a \leq 0$, we speak of a being non-positive

$a \geq 0$, we speak of a being non-negative

Note carefully the distinction between the negative of a ($-a$) and a is negative ($a < 0$).

2. Properties of Inequalities

Like equations, there are certain manipulations that can be performed on inequalities without invalidating the inequality, that is, without changing its sense.

(a) Additive property. If a, b, and c are numbers, and if $a < b$, then

$$a + c < b + c$$

That is, the same quantity may be added to both sides of an inequality without changing its sense.

(b) Multiplicative property. If a, b, and c are numbers, and if $a < b$ and $c > 0$, then

$$a \cdot c < b \cdot c$$

That is, both sides of an inequality may be multiplied by the same positive number without changing its sense.

(c) Transitive property. If a, b, and c are numbers, and $a < b$ and $b < c$, then $a < c$.

(d) If $a > b$, then $a^n > b^n$ if a, b, and n are all positive.

That is, both sides of an inequality of positive numbers may be raised to the same positive power without changing the sense of the inequality.

Notice that the addition property also implies that if $a > b$, then $a - c > b - c$. That is, subtracting equal quantities from both sides of an inequality is equivalent to adding equal negative quantities to both sides.

Also, the multiplicative property implies that if $a > b$, then $\frac{a}{c} > \frac{b}{c}$ if $c > 0$ because dividing both sides by c is equivalent to multiplying both sides by the quantity $\frac{1}{c}$.

The multiplicative property does not remain true for inequalities if we multiply by zero or a negative number. In fact, in the latter case it actually reverses the sense of the inequality. Let's see how.

If $a > b$ and $c < 0$ and we add $-b$ to both sides of $a > b$, we get

$$a - b > 0$$

Then multiplying both sides of this inequality by the positive number $-c$

$$-c\,(a - b) > 0 \cdot (-c)$$

$$bc - ac > 0$$

Adding ac to both sides now, we get

$$bc > ac$$

or

$$ac < bc$$

which has the opposite sense from the original inequality, $a > b$.

Questions

1. Show that $1 > 0$. (Hint: If $1 < 0$, then $-1 > 0$. Remember that $(-1)(-1) = 1$, so $1 > 0$ -- Impossible! Why?)

Discuss and verify for several numbers. (Prove if you can.)

2. If $a > b$, $c > d$, then $a + c > b + d$.
3. If $a > b > 0$ and $c > d > 0$, then $ac > bd$.
4. If $c > 0$ and $a > b$, then $a + c > b$.
5. If $a > b > 0$, then $\frac{1}{a} < \frac{1}{b}$.
6. If $a > 1$, then $a^2 > a$.
7. If $0 < a < 1$, then $a^2 < a$.
8. If $a > 0$, then $\frac{1}{a} > 0$.
9. If $a < 0$, then $\frac{1}{a} < 0$.
10. Given $\frac{s}{c} > x$, $x > s$, $s > 0$, and $1 > c > 0$. Also $c^2 = 1 - s^2$. Show that $1 - x^2 < \frac{s}{x} < 1$ by first showing $\frac{s}{c} > x > s > 0$. Why? Then multiply by $\frac{1}{s}$ and use No. 5 above. Note that $0 < c < 1$ and use No. 7. Then use $x > s$ with this result, and $c^2 = 1 - s^2$ to prove the result.

Appendix 4: Tables

LOGARITHMS OF NUMBERS

N	0	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396

N	0	1	2	3	4	5	6	7	8	9
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996

Table of Trigonometric Functions

sin (read down)

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9		
0°	.0000	.0017	.0035	.0052	.0070	.0087	.0105	.0122	.0140	.0157	.0175	**89°**
1°	.0175	.0192	.0209	.0227	.0244	.0262	.0279	.0297	.0314	.0332	.0349	**88°**
2°	.0349	.0366	.0384	.0401	.0419	.0436	.0454	.0471	.0488	.0506	.0523	**87°**
3°	.0523	.0541	.0558	.0576	.0593	.0610	.0628	.0645	.0663	.0680	.0698	**86°**
4°	.0698	.0715	.0732	.0750	.0767	.0785	.0802	.0819	.0837	.0854	.0872	**85°**
5°	.0872	.0889	.0906	.0924	.0941	.0958	.0976	.0993	.1011	.1028	.1045	**84°**
6°	.1045	.1063	.1080	.1097	.1115	.1132	.1149	.1167	.1184	.1201	.1219	**83°**
7°	.1219	.1236	.1253	.1271	.1288	.1305	.1323	.1340	.1357	.1374	.1392	**82°**
8°	.1392	.1409	.1426	.1444	.1461	.1478	.1495	.1513	.1530	.1547	.1564	**81°**
9°	.1564	.1582	.1599	.1616	.1633	.1650	.1668	.1685	.1702	.1719	.1736	**80°**
10°	.1736	.1754	.1771	.1788	.1805	.1822	.1840	.1857	.1874	.1891	.1908	**79°**
11°	.1908	.1925	.1942	.1959	.1977	.1994	.2011	.2028	.2045	.2062	.2079	**78°**
12°	.2079	.2096	.2113	.2130	.2147	.2164	.2181	.2198	.2115	.2233	.2250	**77°**
13°	.2250	.2267	.2284	.2300	.2317	.2334	.2351	.2368	.2385	.2402	.2419	**76°**
14°	.2419	.2436	.2453	.2470	.2487	.2504	.2521	.2538	.2554	.2571	.2588	**75°**
15°	.2588	.2605	.2622	.2639	.2656	.2672	.2689	.2706	.2723	.2740	.2756	**74°**
16°	.2756	.2773	.2790	.2807	.2823	.2840	.2857	.2874	.2890	.2907	.2924	**73°**
17°	.2924	.2940	.2957	.2974	.2990	.3007	.3024	.3040	.3057	.3074	.3090	**72°**
18°	.3090	.3107	.3123	.3140	.3156	.3173	.3190	.3206	.3223	.3239	.3256	**71°**
19°	.3256	.3272	.3289	.3305	.3322	.3338	.3355	.3371	.3387	.3404	.3420	**70°**
20°	.3420	.3437	.3453	.3469	.3486	.3502	.3518	.3535	.3551	.3567	.3584	**69°**
21°	.3584	.3600	.3616	.3633	.3649	.3665	.3681	.3697	.3714	.3730	.3746	**68°**
22°	.3746	.3762	3778	.3795	.3811	.3827	.3843	.3859	.3875	.3891	.3907	**67°**
23°	.3907	.3923	.3939	.3955	.3971	.3987	.4003	.4019	.4035	.4051	.4067	**66°**
24°	.4067	.4083	.4099	.4115	.4131	.4147	.4163	.4179	.4195	.4210	.4226	**65°**
25°	.4226	.4242	.4258	.4274	.4289	.4305	.4321	.4337	.4352	.4368	.4384	**64°**
26°	.4384	.4399	.4415	.4431	.4446	.4462	.4478	.4493	.4509	.4524	.4540	**63°**
27°	.4540	.4555	.4571	.4586	.4602	.4617	.4633	.4648	.4664	.4679	.4695	**62°**
28°	.4695	.4710	.4726	.4741	.4756	.4772	.4787	.4802	.4818	.4833	.4848	**61°**
29°	.4848	.4863	.4879	.4894	.4909	.4924	.4939	.4955	.4970	.4985	.5000	**60°**
30°	.5000	.5015	.5030	.5045	.5060	.5075	.5090	.5105	.5120	.5135	.5150	**59°**
31°	.5150	.5165	.5180	.5195	.5210	.5225	.5240	.5255	.5270	.5284	.5299	**58°**
32°	.5299	.5314	.5329	.5344	.5358	.5373	.5388	.5402	.5417	.5432	.5446	**57°**
33°	.5446	.5461	.5476	.5490	.5505	.5519	.5534	.5548	.5563	.5577	.5592	**56°**
34°	.5592	.5606	.5621	.5635	.5650	.5664	.5678	.5693	.5707	.5721	.5736	**55°**
35°	.5736	.5750	.5764	.5779	.5793	.5807	.5821	.5835	.5850	.5864	.5878	**54°**
36°	.5878	.5892	.5906	.5920	.5934	.5948	.5962	.5976	.5990	.6004	.6018	**53°**
37°	.6018	.6032	.6046	.6060	.6074	.6088	.6101	.6115	.6129	.6143	.6157	**52°**
38°	.6157	.6170	.6184	.6198	.6211	.6225	.6239	.6252	.6266	.6280	.6293	**51°**
39°	.6293	.6307	.6320	.6334	.6347	.6361	.6374	.6388	.6401	.6414	.6428	**50°**
40°	.6428	.6441	.6455	.6468	.6481	.6494	.6508	.6521	.6534	.6547	.6561	**49°**
41°	.6561	.6574	.6587	.6600	.6613	.6626	.6639	.6652	.6665	.6678	.6691	**48°**
42°	.6691	.6704	.6717	.6730	.6743	.6756	.6769	.6782	.6794	.6807	.6820	**47°**
43°	.6820	.6833	.6845	.6858	.6871	.6884	.6896	.6909	.6921	.6934	.6947	**46°**
44°	.6947	.6959	.6972	.6984	.6997	.7009	.7022	.7034	.7046	.7059	.7071	**45°**
		.9	.8	.7	.6	.5	.4	.3	.2	.1	.0	

cos (read up)

Table of Trigonometric Functions

sin (read down)

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9		
45°	.7071	.7083	.7096	.7108	.7120	.7133	.7145	.7157	.7169	.7181	.7193	**44°**
46°	.7193	.7206	.7218	.7230	.7242	.7254	.7266	.7278	.7290	.7302	.7314	**43°**
47°	.7314	.7325	.7337	.7349	.7361	.7373	.7385	.7396	.7408	.7420	.7431	**42°**
48°	.7431	.7443	.7455	.7466	.7478	.7490	.7501	.7513	.7524	.7536	.7547	**41°**
49°	.7547	.7559	.7570	.7581	.7593	.7604	.7615	.7627	.7638	.7649	.7660	**40°**
50°	.7660	.7672	.7683	.7694	.7705	.7716	.7727	.7738	.7749	.7760	.7771	**39°**
51°	.7771	.7782	.7793	.7804	.7815	.7826	.7837	.7848	.7859	.7869	.7880	**38°**
52°	.7880	.7891	.7902	.7912	.7923	.7934	.7944	.7955	.7965	.7976	.7986	**37°**
53°	.7986	.7997	.8007	.8018	.8028	.8039	.8049	.8059	.8070	.8080	.8090	**36°**
54°	.8090	.8100	.8111	.8121	.8131	.8141	.8151	.8161	.8171	.8181	.8192	**35°**
55°	.8192	.8202	.8211	.8221	.8231	.8241	.8251	.8261	.8271	.8281	.8290	**34°**
56°	.8290	.8300	.8310	.8320	.8329	.8339	.8348	.8358	.8368	.8377	.8387	**33°**
57°	.8387	.8396	.8406	.8415	.8425	.8434	.8443	.8453	.8462	.8471	.8480	**32°**
58°	.8480	.8490	.8499	.8508	.8517	.8526	.8536	.8545	.8554	.8563	.8572	**31°**
59°	.8572	.8581	.8590	.8599	.8607	.8616	.8625	.8634	.8643	.8652	.8660	**30°**
60°	.8660	.8669	.8678	.8686	.8695	.8704	.8712	.8721	.8729	.8738	.8746	**29°**
61°	.8746	.8755	.8763	.8771	.8780	.8788	.8796	.8805	.8813	.8821	.8829	**28°**
62°	.8829	.8838	.8846	.8854	.8862	.8870	.8878	.8886	.8894	.8902	.8910	**27°**
63°	.8910	.8918	.8926	.8934	.8942	.8949	.8957	.8965	.8973	.8980	.8988	**26°**
64°	.8988	.8996	.9003	.9011	.9018	.9026	.9033	.9041	.9048	.9056	.9063	**25°**
65°	.9063	.9070	.9078	.9085	.9092	.9100	.9107	.9114	.9121	.9128	.9135	**24°**
66°	.9135	.9143	.9150	.9157	.9164	.9171	.9178	.9184	.9191	.9198	.9205	**23°**
67°	.9205	.9212	.9219	.9225	.9232	.9239	.9245	.9252	.9259	.9265	.9272	**22°**
68°	.9272	.9278	.9285	.9291	.9298	.9304	.9311	.9317	.9323	.9330	.9336	**21°**
69°	.9336	.9342	.9348	.9354	.9361	.9367	.9373	.9379	.9385	.9391	.9397	**20°**
70°	.9397	.9403	.9409	.9415	.9421	.9426	.9432	.9438	.9444	.9449	.9455	**19°**
71°	.9455	.9461	.9466	.9472	.9478	.9483	.9489	.9494	.9500	.9505	.9511	**18°**
72°	.9511	.9516	.9521	.9527	.9532	.9537	.9542	.9548	.9553	.9558	.9563	**17°**
73°	.9563	.9568	.9573	.9578	.9583	.9588	.9593	.9598	.9603	.9608	.9613	**16°**
74°	.9613	.9617	.9622	.9627	.9632	.9636	.9641	.9646	.9650	.9655	.9659	**15°**
75°	.9659	.9664	.9668	.9673	.9677	.9681	.9686	.9690	.9694	.9699	.9703	**14°**
76°	.9703	.9707	.9711	.9715	.9720	.9724	.9728	.9732	.9736	.9740	.9744	**13°**
77°	.9744	.9748	.9751	.9755	.9759	.9763	.9767	.9770	.9774	.9778	.9781	**12°**
78°	.9781	.9785	.9789	.9792	.9796	.9799	.9803	.9806	.9810	.9813	.9816	**11°**
79°	.9816	.9820	.9823	.9826	.9829	.9833	.9836	.9839	.9842	.9845	.9848	**10°**
80°	.9848	.9851	.9854	.9857	.9860	.9863	.9866	.9869	.9871	.9874	.9877	**9°**
81°	.9877	.9880	.9882	.9885	.9888	.9890	.9893	.9895	.9898	.9900	.9903	**8°**
82°	.9903	.9905	.9907	.9910	.9912	.9914	.9917	.9919	.9921	.9923	.9925	**7°**
83°	.9925	.9928	.9930	.9932	.9934	.9936	.9938	.9940	.9942	.9943	.9945	**6°**
84°	.9945	.9947	.9949	.9951	.9952	.9954	.9956	.9957	.9959	.9960	.9962	**5°**
85°	.9962	.9963	.9965	.9966	.9968	.9969	.9971	.9972	.9973	.9974	.9976	**4°**
86°	.9976	.9977	.9978	.9979	.9980	.9981	.9982	.9983	.9984	.9985	.9986	**3°**
87°	.9986	.9987	.9988	.9989	.9990	.9990	.9991	.9992	.9993	.9993	.9994	**2°**
88°	.9994	.9995	.9995	.9996	.9996	.9997	.9997	.9997	.9998	.9998	.9998	**1°**
89°	.9998	.9999	.9999	.9999	.9999	1.000	1.000	1.000	1.000	1.000	1.000	**0°**
		.9	.8	.7	.6	.5	.4	.3	.2	.1	.0	

cos (read up)

TRIGONOMETRIC FUNCTIONS FOR ANGLES IN RADIANS

Rad	Sin	Cos	Rad.	Sin	Cos	Rad.	Sin	Cos	Rad.	Sin	Cos
.00	.00000	1.00000	**.40**	.38942	.92106	**.80**	.71736	.69671	**1.20**	.93204	.36236
.01	.10000	0.99995	.41	.39861	.91712	.81	.72429	.68950	1.21	.93562	.35302
.02	.02000	.99980	.42	.40776	.91309	.82	.73115	.68222	1.22	.93910	.34365
.03	.03000	.99955	.43	.41687	.90897	.83	.73793	.67488	1.23	.94249	.33424
.04	.03999	.99920	.44	.42594	.90475	.84	.74464	.66746	1.24	.94578	.32480
.05	.04998	.99875	.45	.43497	.90045	.85	.75128	.65998	1.25	.94898	.31532
.06	.05996	.99820	.46	.44395	.89605	.86	.75784	.65244	1.26	.95209	.30582
.07	.06994	.99755	.47	.45289	.89157	.87	.76433	.64483	1.27	.95510	.29628
.08	.07991	.99680	.48	.46178	.88699	.88	.77074	.63715	1.28	.95802	.28672
.09	.08988	.99595	.49	.47063	.88233	.89	.77707	.62941	1.29	.96084	.27712
.10	.09983	.99500	**.50**	.47943	.87758	**.90**	.78333	.62161	**1.30**	.96356	.26750
.11	.10978	.99396	.51	.48818	.87274	.91	.78950	.61375	1.31	.96618	.25785
.12	.11971	.99281	.52	.49688	.86782	.92	.79560	.60582	1.32	.96872	.24818
.13	.12963	.99156	.53	.50553	.86281	.93	.80162	.59783	1.33	.97115	.23848
.14	.13954	.99022	.54	.51414	.85771	.94	.80756	.58979	1.34	.97348	.22875
.15	.14944	.98877	.55	.52269	.85252	.95	.81342	.58168	1.35	.97572	.21901
.16	.15932	.98723	.56	.53119	.84726	.96	.81919	.57352	1.36	.97786	.20924
.17	.16918	.98558	.57	.53963	.84190	.97	.82489	.56530	1.37	.97991	.19945
.18	.17903	.98384	.58	.54802	.83646	.98	.83050	.55702	1.38	.98185	.18964
.19	.18886	.98200	.59	.55636	.83094	.99	.83603	.54869	1.39	.98370	.17981
.20	.19867	.98007	**.60**	.56464	.82534	**1.00**	.84147	.54030	**1.40**	.98545	.16997
.21	.20846	.97803	.61	.57287	.81965	1.01	.84683	.53186	1.41	.98710	.16010
.22	.21823	.97590	.62	.58104	.81388	1.02	.85211	.52337	1.42	.98865	.15023
.23	.22798	.97367	.63	.58914	.80803	1.03	.85730	.51482	1.43	.99010	.14033
.24	.23770	.97134	.64	.59720	.80210	1.04	.86240	.50622	1.44	.99146	.13042
.25	.24740	.96891	.65	.60519	.79608	1.05	.86742	.49757	1.45	.99271	.12050
.26	.25708	.96639	.66	.61312	.78999	1.06	.87236	.48887	1.46	.99387	.11057
.27	.26673	.96377	.67	.62099	.78382	1.07	.87720	.48012	1.47	.99492	.10063
.28	.27636	.96106	.68	.62879	.77757	1.08	.88196	.47133	1.48	.99588	.09067
.29	.28595	.95824	.69	.63654	.77125	1.09	.88663	.46249	1.49	.99674	.08071
.30	.29552	.95534	**.70**	.64422	.76484	**1.10**	.89121	.45360	**1.50**	.99749	.07074
.31	.30506	.95233	.71	.65183	.75836	1.11	.89570	.44466	1.51	.99815	.06076
.32	.31457	.94924	.72	.65938	.75181	1.12	.90010	.43568	1.52	.99871	.05077
.33	.32404	.94604	.73	.66687	.74517	1.13	.90441	.42666	1.53	.99917	.04079
.34	.33349	.94275	.74	.67429	.73847	1.14	.90863	.41759	1.54	.99953	.03079
.35	.34290	.93937	.75	.68164	.73169	1.15	.91276	.40849	1.55	.99978	.02079
.36	.35227	.93590	.76	.68892	.72484	1.16	.91680	.39934	1.56	.99994	.01080
.37	.36162	.93233	.77	.69614	.71791	1.17	.92075	.39015	1.57	1.00000	.00080
.38	.37092	.92866	.78	.70328	.71091	1.18	.92461	.38092	1.58	.99996	-.00920
.39	.38019	.92491	.79	.71035	.70385	1.19	.92837	.37166	1.59	.99982	-.01920